

Biblioteconomia e scienza dell'informazione

**MAURO GUERRINI
TIZIANA POSSEMATO**

LINKED DATA PER BIBLIOTECHE, ARCHIVI E MUSEI



EDITRICE BIBLIOGRAFICA

**BIBLIOTECONOMIA
E SCIENZA DELL'INFORMAZIONE**

8.

Mauro Guerrini – Tiziana Possemato

Linked data per biblioteche, archivi e musei

*Perché l'informazione sia del web
e non solo nel web*

Con un saggio di Carlo Bianchini
e la consulenza di Rosa Maiello e Valdo Pasqui

Prefazione di Roberto delle Donne

EDITRICE BIBLIOGRAFICA

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Le fotocopie effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali, corso di Porta Romana n. 108, 20122 Milano, e-mail: autorizzazioni@clearedi.org e sito web: www.clearedi.org.

Copertina: MoskitoDesign - Varese
Fotocomposizione:
Arti Grafiche Colombo - Muggiò (MB)
ISBN: 978-88-7075-852-8
Copyright © 2015 Editrice Bibliografica
Via F. De Sanctis, 33/35 - 20141 Milano
Proprietà letteraria riservata - Printed in Italy

A p. Silvano Danieli
persona capace di
connettere persone

Ανθρωπος φυσει πολιτικον ζων.
Aristotele

In an extreme view, the world can be seen as only connections, nothing else. We think of a dictionary as the repository of meaning, but it defines words only in terms of other words. I liked the idea that a piece of information is really defined only by what it's related to, and how it's related. There really is little else to meaning. The structure is everything.

Tim Berners-Lee

This changes everything. Again.
Steve Jobs

PREMESSA

Ἄνθρωπος φύσει πολιτικὸν ζῷον, scrive Aristotele in Τὰ πολιτικά (*La politica*); l'uomo è un animale politico, un animale sociale, ed è portato per sua natura (φύσει) a cercare la compagnia dei propri simili per creare relazioni e per formare comunità. Alla base di ciò si trova il λόγος: ogni relazione e ogni scambio avviene tramite un linguaggio. Il bisogno di comunicare ha accompagnato l'uomo nell'arco di tutta la sua evoluzione; si può addirittura ipotizzare che la storia della specie umana sarebbe giunta al termine migliaia di anni fa se non avesse fatto uso di metodi per comunicare e tramandare le conoscenze acquisite: dal linguaggio dei segni, alle pitture rupestri, alla parola, alla scrittura.¹

I linguaggi per la trasmissione della conoscenza registrata si evolvono in continuazione sulla base della tradizione culturale e linguistica acquisita, delle caratteristiche sempre più variegata dell'universo bibliografico, della disponibilità di tecnologie più efficaci e delle esigenze più puntuali delle persone. I linked data rappresentano l'ultima tappa di tale processo.

Questo libro riflette sui principi e sui meccanismi dei linked data (dati connessi), un nuovo linguaggio della comunicazione globale concepito a cavallo tra XX e XXI secolo, e sul valore che i dati connessi assumono specificatamente nel contesto della comunicazione. La tematica è analizzata nella dimensione storica, teorica, tecnica e funzionale. I linked data possono essere creati e utilizzati per qualsiasi tipo di risorsa e per qualsiasi finalità in ambito bibliotecario, archivistico e museale, così come in ambito amministrativo, legislativo, commerciale, medico e della ricerca; essi possono, infatti, essere usati da tutte le agenzie che creano i dati, affinché l'informazione prodotta sia *del* web e non solo *nel* web, come ha autorevolmente affermato Karen Coyle in varie occasioni.

Il web è il luogo in cui ormai si manifestano e si svolgono molti aspetti della vita professionale e privata delle persone; esso

¹ La suggestione è di Helena Bucci, studentessa di biblioteconomia, Università di Firenze, a.a. 2014-2015.

ha rivoluzionato il modo di stabilire le relazioni e di fruire le informazioni disponibili e può rivoluzionare il modo di *scoprire*, *accedere*, *integrare* e *utilizzare* i dati per le sue caratteristiche di *ubiquità* (virtuale) e per la sua *natura distribuita e scalabile* (modulare). L'assunto è alla base della riflessione sulle modalità e sulle strategie di diffusione dell'informazione nella nostra epoca.

Nell'ottobre 2011, il Library Linked Data Incubator Group (LLD XG), creato presso il World Wide Web Consortium (W3C), ha pubblicato una relazione in cui sosteneva la necessità che le biblioteche, i servizi d'informazione bibliografica (in particolare le agenzie bibliografiche nazionali), gli archivi e i musei rendessero disponibili i loro dati nel web come linked data,² ovvero in una *forma significativa*, vale a dire in una forma adatta a consentire la comprensione del loro significato anche da parte delle applicazioni informatiche. La tecnologia usata avrebbe dovuto basarsi su metadati estremamente strutturati che consentissero alle macchine di riconoscere le relazioni tra le entità e di applicare processi inferenziali; ciò rappresenta un'opportunità di soluzione a molte criticità comuni alle istituzioni della memoria registrata, come la moltiplicazione dei dati bibliografici nel web e l'identificazione esatta delle entità (per esempio, nel caso di omonimia o di cambio di nome per un'entità di tipo *Persona*).

Il volume presenta i linguaggi di metadatazione, le modalità di costruzione delle triple e l'uso delle ontologie; introduce e descrive gli standard, le tecnologie e i linguaggi per la creazione, pubblicazione, utilizzo, gestione e ciclo di vita dei linked data (come formulare, collegare e pubblicare i metadati nel web), in modo da consentire al lettore di comprendere e interpretare la loro natura e la loro funzione. Offre una panoramica delle pratiche di applicazione più interessanti realizzate da importanti istituzioni, prevalentemente in ambito bibliotecario. Discute delle novità teoriche e applicative in ambito catalografico, l'effetto di FRBR sui modelli descrittivi, il rapporto tra RDA e la tecnologia dei linked data; la relazione con MARC, le nuove prospettive

² Per una presentazione e un commento del rapporto, vedi: Seth Van Hooland, Ruben Verborgh, *Linked data for libraries, archives and museums. How to clean, link and publish your metadata*. London: Facet Publishing, 2014.

aperte da BIBFRAME; esamina le problematiche relative alla gestione dei diritti di distribuzione e d'uso dei linked data; presenta un glossario essenziale dei concetti principali che ritornano costantemente nel volume. Il passaggio dal record al dato genera una trasformazione profonda negli strumenti di mediazione tra universo bibliografico e utente che nessun attore che operi in questo ambito può ignorare. Per questi aspetti, *Linked data per biblioteche, archivi e musei* si presenta come l'altro lato, ovvero come la prosecuzione naturale di *Introduzione a RDA*,³ in quanto i linked data esprimono la dimensione tecnologica necessaria per concretizzare al meglio RDA, il primo standard di metadattazione concepito per la creazione dei dati e la scoperta delle risorse nell'era digitale; scopo di RDA è, infatti, proporre linee guida per *identificare* e *collegare* le risorse. RDA e linked data rappresentano, pertanto, un binomio naturale: le linee guida e la tecnologia sono state pensate per gli ILS di nuova generazione.

Il percorso di elaborazione del libro

Gli autori hanno iniziato il percorso di analisi e di studio nel 2010; il primo risultato è stato la pubblicazione del saggio *Linked data: un nuovo alfabeto del web semantico*, in "Biblioteche oggi", vol. 30, n. 3 (aprile 2012), p. 7-15,⁴ presentato originariamente come relazione al convegno *I nuovi alfabeti della biblioteca*, Milano, Palazzo delle Stelline, 15-16 marzo 2012; il cammino è proseguito con la preparazione del seminario internazionale *Global interoperability and linked data in libraries*, Firenze 18-19 giugno 2012, promosso da numerosi istituti: Università di Firenze, Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), Biblioteca nazionale centrale di Firenze (BNCF), Casalini Libri, Comune di Firenze, Conferenza dei rettori delle università italiane (CRUI), Istituto di teoria e tecniche dell'informazione giuridica del

³ Carlo Bianchini, Mauro Guerrini, *Introduzione a RDA. Linee guida per rappresentare e scoprire le risorse*, prefazione di Barbara B. Tillett; postfazione di Gordon Dunsire. Milano: Editrice Bibliografica, 2014.

⁴ <www.bibliotecheoggi.it/content/201200300701.pdf>.

Consiglio nazionale delle ricerche (ITTIG-CNR), Fondazione Rinascimento digitale e Associazione italiana biblioteche (AIB), con la sponsorizzazione di OCLC e @Cult; le relazioni sono state pubblicate su “Jlis.it”, vol. 4, no. 1 (gennaio/January 2012), disponibile open access⁵ e, in formato cartaceo, on demand, da Ledizioni di Milano. Il 18 aprile 2013 Tiziana Possemato ha discusso la tesi di Master di secondo livello in Biblioteconomia, all’Università di Firenze, dal titolo *Del web e non solo nel web. Le risorse nel web semantico*, relatore il prof. Mauro Guerrini, correlatore prof. Graziano Ruffini. La tematica è stata presentata in diversi incontri e seminari, in particolare, al corso promosso dall’AIB nazionale all’Università di Napoli Federico II il 28-29 maggio 2012, dall’AIB Trentino-Alto Adige alla Libera Università di Bolzano il 16-17 settembre 2013, dall’AIB Sicilia all’Università di Palermo il 7-8 novembre 2013 quale iniziativa del MAB Sicilia, e dalla Sapienza Università di Roma il 30-31 ottobre 2014, con il supporto dell’AIB Lazio. È stato motivo di lezione al Master di secondo livello in Biblioteconomia dell’Università di Firenze e al dottorato in Studi storici delle Università di Firenze e Siena.

Il libro si rivolge agli studenti di corsi avanzati di biblioteconomia, archivistica e scienze dell’informazione, ai bibliotecari, agli archivisti, ai conservatori di museo, alle agenzie che creano i dati, ai professionisti e a tutti coloro che si occupano di strutturare l’informazione per la diffusione nel web.

Firenze – Roma, 26 gennaio 2015

⁵ <<http://leo.cineca.it/index.php/jlis/issue/view/536>>.

Ringraziamenti

Graziano Ruffini ha affettuosamente seguito il concepimento e la progressione del libro, dalla discussione della tesi di master di Tiziana come correlatore, alle discussioni su singoli temi avvenute nel corso degli anni, fino alla redazione finale, ascoltando con pazienza certi entusiasmi iniziali e fornendo suggerimenti nel corso della stesura del testo. Valdo Pasqui ha fornito un supporto indispensabile per la definizione di concetti e procedure; ha letto accuratamente il testo, intervenendo autorevolmente con proposte di modifica e chiarimento concettuale in numerose sue parti; ha formulato numerose definizioni del glossario; ha redatto lo schema del paragrafo 2.6.2.

Rosa Maiello ha proposto miglioramenti stilistici e ha contribuito con grande competenza a redigere i capitoli 6 e 7 sulla gestione dei diritti di pubblicazione e sulla normativa relativa ai linked data.

Elena Escolano Rodríguez ha letto con disponibilità e amicizia il testo finale, proponendo soluzioni migliorative che derivano dalla sua grandissima competenza teorica e pratica in ambito internazionale.

Iryna Solodovnik ha letto meticolosamente il testo proponendo modifiche puntuali, redatto il paragrafo 2.7.2 e l'appendice *Vocabolario o Schema LOD-ready per l'interlinking con metadati Responsible Body e Subject e per la loro abilitazione come LOD*; ha rivisto diverse voci del glossario, grazie alla profonda conoscenza del tema, trattato nella sua tesi di dottorato di ricerca *Strategie linked data per gli archivi istituzionali open access. Per una migliore comunicazione dei contenuti di ricerca*, Università della Calabria, a.a. 2012-2013, di prossima pubblicazione, in una forma rivista, per la Firenze University Press (FUP).

Giovanna Aracri ha letto il testo in una versione iniziale e in una versione semidefinitiva, suggerendo modifiche pertinenti, in quanto i linked data rientrano tra i suoi principali interessi di ricerca; ha in corso, infatti, la discussione della tesi di dottorato di ricerca *Interoperabilità e semantica nei domini specialistici: l'approccio linked data nel settore delle costruzioni*, al dottorato in Scienze bibliografiche, del testo e del documento (XXVII ciclo) all'Università di Udine, relatore Mauro Guerrini.

Alberto Salarelli ha letto scrupolosamente il testo e ha suggerito proposte di modifica.

Carla Martini ha seguito attentamente l'evolversi del lavoro nella fase finale, proposto miglioramenti nella formulazione di frasi del testo e del glossario.

Andrea Gazzarini, ha redatto il paragrafo 8.7 dedicato al Progetto europeo ALIADA e ha rivisto gli schemi del capitolo 4.

Elisabetta Viti ha fornito la definizione della voce *Thesaurus del Nuovo Soggettario in formato SKOS/RDF*, pubblicata nell'Appendice *Vocabolario o Schema LOD-ready*.

Emanuele Bellini, Eleonora Belpassi, Gianfranco Crupi, Klaus Kempft, Claudio Leombroni, Mario Mancini, Andrea Marchitelli, Patrizia Martini, Ginevra Peruginelli, Stefano Zamponi hanno letto e commentato alcune parti del volume. L'inizio della *Premessa* cerca di rispondere a un'obiezione, o, più esattamente, a una richiesta di chiarimento di Stefano circa l'innesto della tematica dei linked data nella tradizione degli studi umanistici.

Daniela Giglio ha letto il testo in varie versioni e ha redatto l'indice dei nomi.

Ivo Contursi ha favorito, con grande generosità, la redazione del volume, mettendo a disposizione degli autori la competenza professionale e scientifica di @Cult, un'azienda leader nell'applicazione dei linked data a progetti per biblioteche e istituti culturali.

PREFAZIONE

di *Roberto Delle Donne*

Questo libro di Mauro Guerrini e di Tiziana Possemato colma una lacuna e risponde a un'attesa vivamente sentita non solo dagli studiosi di biblioteconomia, archivistica e museografia, ai quali esplicitamente si rivolge, ma anche dai molti ricercatori operanti nei più diversi ambiti disciplinari attenti al modo in cui le reti telematiche hanno trasformato le pratiche della ricerca scientifica, le forme di comunicazione del sapere e le modalità di diffusione dei suoi risultati. Nell'ultimo decennio, con il moltiplicarsi dei produttori di contenuti in grado di inserire facilmente in internet informazione strutturata, è infatti cresciuta l'esigenza di descrivere e organizzare tale informazione, per garantirne la reperibilità ed evitare che si perda nel rumore di fondo della rete. In tale prospettiva sono stati determinanti gli orientamenti del web semantico o *web di dati*, secondo cui il lavoro di organizzazione e di gestione dell'informazione dev'essere in gran parte automatico e basato su descrizioni fortemente standardizzate e formalizzate, elaborate da specialisti. I sistemi di classificazione dell'informazione alla base del web semantico sono ontologie formali, schemi di ordinamento dei documenti gerarchico-enumerativi oppure analitico-sintetici elaborati da esperti del settore, espressi in maniera uniforme e rigorosa e associati all'informazione primaria attraverso l'uso di linguaggi e formalismi, a loro volta rigidamente strutturali e ben definiti, comprensibili dalle macchine. In questo contesto, al web dei documenti basato sul linguaggio HTML si aggiunge un insieme strutturato di informazioni collegate tra loro: i cosiddetti *linked data*, i dati connessi o dati collegati. Essi costituiscono, quindi, una tecnologia e un insieme di buone pratiche per pubblicare dati sul web in una modalità leggibile, interpretabile e omogeneizzabile da agenti non umani, come, per esempio, i motori di ricerca.

Di *linked data* Mauro Guerrini si è già occupato, con il rigore e la competenza che lo caratterizzano, in un numero monografico della rivista JLIS.it da lui curato (no. 4, 1 del 2013), nonché in un recente volume pubblicato insieme a Carlo Bianchini, dedicato a RDA (*Resource Description and Access*), uno standard

internazionale utilizzabile da chiunque voglia descrivere e rendere accessibili in rete risorse culturali di qualsiasi tipo, favorendo anche l'integrazione dei cataloghi delle biblioteche con gli altri strumenti di informazione.¹ RDA mostra, infatti, come costruire metadati di qualità che possano essere riutilizzati per scopi molteplici, con la massima efficacia, innanzitutto in un contesto di linked data. Tiziana Possemato è molto nota in Italia per avere esplorato caratteri e potenzialità di questa tecnologia anche attraverso la realizzazione di *Open Linked Data Framework*, una piattaforma per la creazione e per la pubblicazione di linked data, rilasciata con il nome *Quickly*, già adottata dal Museo Galileo di Firenze, dalla Provincia autonoma di Trento per una sperimentazione sui dati culturali, nonché da una rete di atenei italiani confederati in vista della condivisione dei cataloghi (Università di Napoli Federico II, Napoli L'Orientale, Napoli Parthenope, Salerno, Sannio, Basilicata, Salento).

D'altronde, da alcuni anni si parla molto di *open data*, di *linked data* e di *open linked data* in diversi ambiti. Mi limito a menzionarne solo alcuni. Numerose amministrazioni pubbliche, per impulso della Comunità Europea (Agenda digitale europea, Action 3), hanno scelto di rendere liberamente accessibile sul web come open data l'enorme patrimonio di informazioni che raccolgono e detengono in virtù dei propri ruoli istituzionali. In Europa, alcune amministrazioni pubbliche hanno anche identificato nelle tecnologie standard del web semantico, e in particolare nel modello degli open linked data, gli strumenti imprescindibili per dare ai dati (aperti o non) un'identità, per renderli collegabili e semanticamente interoperabili tra loro, secondo quanto auspicato dall'Agenda digitale europea.

Un forte impulso alla condivisione e all'interoperabilità dei dati è presente anche nel contesto della ricerca scientifica, dal momento che la Commissione Europea, la Max-Planck-Gesellschaft e il britannico HM Treasury, con le loro dichiarazioni e raccomandazioni, con le loro policy e i loro programmi di finanziamento come Horizon 2020, incoraggiano gli studiosi a considerare unitariamente il ciclo della ricerca e a renderlo pubblico in rete nella sua interezza, secondo standard internazionali:

¹ Carlo Bianchini, Mauro Guerrini, *Introduzione a RDA*, cit.

dalla raccolta alla classificazione dei dati, dalla loro strutturazione relazionale alla loro presentazione testuale, fino alla loro rielaborazione nell'ordine discorsivo di un articolo o di un volume. In molte comunità disciplinari, anche dell'area umanistica, si è, quindi, avviata da tempo un'attenta riflessione sui dataset e sui caratteri che i repository dei dati dovrebbero avere, sul modo in cui ne andrebbe assicurata l'accessibilità, l'interoperabilità e la conservazione nel lungo periodo, su come renderli citabili e riutilizzabili da altri, sulle peculiari forme che dovrebbe assumere la loro peer review.

Il libro di Mauro Guerrini e di Tiziana Possemato sui principi fondamentali dei linked data rappresenta, quindi, un importante contributo di analisi e di riflessione storica, teorica, tecnica e funzionale su un tema di estrema attualità e di enorme importanza per il futuro, dal momento che delinea modalità e strategie di integrazione di fonti informative eterogenee in un'unica piattaforma di ricerca ed è sorretto dalla volontà di promuovere la cultura dell'apertura dei saperi e della scienza.

1. WEB SEMANTICO E LINKED DATA: DEFINIZIONI E PRINCIPI

1.1 Ripensare le strategie dell'informazione

“Ho fatto un sogno riguardante il Web ... ed è un sogno diviso in due parti. Nella prima parte, il Web diventa il mezzo di gran lunga più potente per favorire la collaborazione tra i popoli. Ho sempre immaginato lo spazio dell'informazione come una cosa a cui tutti abbiano accesso immediato e intuitivo, non solo per navigare ma anche per creare [dati. ...]. Inoltre, il sogno della comunicazione diretta tramite il sapere condiviso dev'essere possibile per gruppi di qualsiasi dimensione, gruppi che potranno interagire elettronicamente con la medesima facilità che facendolo di persona. Nella seconda parte del sogno, la collaborazione si allarga ai computer. Le macchine diventano capaci di analizzare tutti i dati nel web, il contenuto, i link e le transazioni tra persone e computer. La “Rete Semantica” che dovrebbe renderlo possibile deve ancora nascere, ma quando l'avremo i meccanismi quotidiani di commercio, burocrazia e vita saranno gestiti da macchine che parleranno a macchine, lasciando che gli uomini pensino soltanto a fornire l'ispirazione e l'intuito. Finalmente, si materializzeranno quegli “agenti” intelligenti sognati per decenni. Questo Web comprensibile alle macchine si concretizzerà introducendo una serie di progressi tecnici e di adeguamenti sociali attualmente in fase di sviluppo [...]. Una volta realizzato questo duplice sogno, il Web sarà un luogo in cui l'improvvisazione dell'essere umano e il ragionamento della macchina coesisteranno in una miscela ideale e potente”.¹

È il sogno di Tim Berners-Lee, che nel marzo 1989, da scienziato al CERN (European Organization for Nuclear Research,

¹ Tim Berners-Lee, *L'architettura del nuovo web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*; in collaborazione con Mark Fischetti. Milano: Feltrinelli, 2001. Traduzione di: *Weaving the web: the past, present and future of the world wide web by its inventor*; with Mark Fischetti. London: Orion Business Book, 1999, p. 139-140.

Organizzazione europea per la ricerca nucleare) di Ginevra, concepì il world wide web (WWW): egli ipotizzò e realizzò un mezzo per condividere i documenti tra i ricercatori delle università, dei centri di ricerca e degli istituti di ogni paese. La soluzione fu escogitata per rispondere alla domanda di condivisione delle informazioni tra gli utenti sparsi nel mondo, utilizzando internet: creare la rete delle reti, unire le reti di computer e il meccanismo dell'ipertesto in un sistema d'informazione globale potente e semplice. Nel 2001 Berners-Lee propose un'estensione del web che avrebbe permesso alle persone di sfruttare i computer per trattare le informazioni in una modalità più efficace. Chiamò la soluzione *web semantico*, perché questo web avrebbe aggiunto un'infrastruttura con significato concettuale definito – la *semantica* – al di sopra dell'infrastruttura sintattica di HTML (HyperText Markup Language) per il trattamento automatico dei documenti nel web.² Web semantico stabiliva una nuova forma di web, in cui i dati avrebbero potuto essere processati dalle macchine; si trattava di un'evoluzione del web tradizionale. Il gruppo di ricerca sul web semantico del World Wide Web Consortium definisce un framework (una cornice, un ambito) comune in cui condividere e riutilizzare i dati, superando i confini tra applicazioni, enti e comunità differenti. Gli elementi fondamentali di novità erano:

- collegamenti tra *dati* e non tra documenti;
- collegamenti *tipizzati, qualificati*.

Il web semantico diventa un *movimento di collaborazione*, guidato dal World Wide Web Consortium, organismo internazionale di normalizzazione. Il W3C sviluppa tecnologie che garantiscono l'interoperabilità (specifiche, linee guida, software e applicazioni) per portare il World Wide Web al massimo del suo potenziale, agendo da forum di informazioni, comunicazioni e attività comuni. Il W3C italiano (W3C-IT) è il punto di contatto per le attività del W3C in Italia.

Il web semantico si propone di convertire il web attuale, ca-

² Vedi: Mirna Willer, Gordon Dunsire, *Bibliographic information organization in the semantic web*. Oxford, Cambridge; New Delhi: Chandos Publishing, 2013, p. 97-98.

ratterizzato da documenti non strutturati e semi-strutturati, in un web di dati, incoraggiando l'inserimento di contenuto semantico nelle pagine web. L'idea di web semantico può essere affrontata anzitutto come un *fenomeno linguistico*: il metodo che permette l'integrazione coerente di dati diversi. Esso si presenta come una *lingua* per i dati, di una lingua inventata dagli umani per *comunicare informazioni* e pensieri fondamentalmente umani e per l'uso umano, una lingua, tuttavia, concepita per essere *letta ed elaborata* non da lettori o ascoltatori umani, bensì *da computer* secondo algoritmi ben definiti.³

1.2 Il web di documenti e il web di dati

Le due generazioni di web possono essere rappresentate tramite due prospetti presentati alla 7th International World Wide Web Conference W3C Track @ WWW2008, Beijing, China 23-24 April 2008 - *Linked Data: Principles and State of the Art*, riprodotti ai paragrafi successivi.⁴

1.2.1 Il web di documenti

I componenti del web di documenti (o web ipertestuale) sono:

- HTML: linguaggio di marcatura che consente di specificare informazioni sulla modalità di *visualizzazione* del documento nel web;
- collegamenti non tipizzati (o non qualificati; untyped links): tra documenti; indicano che due documenti sono legati, relazionati, ma lasciano all'utente il compito di *desumere la natura della relazione*;
- web-API: una web API (Application Program Interface);
- database (A, B, C, ...): contenitori entro i quali sono conservate le informazioni poi passate al web.

³ Vedi: Tomas Baker, *Designing data for the open world of the web*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/6308/7866>>, p. 1.

⁴ <<http://www.w3.org/2008/Talks/WWW2008-W3CTrack-LOD.pdf>>.

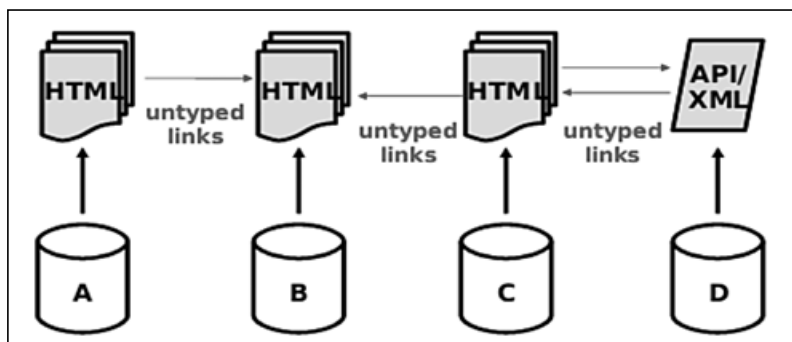


Figura 1 – Rappresentazione del web di documenti

Il *web di documenti* è il web in cui i testi sono pubblicati in pagine HTML; il linguaggio con cui vengono scritti i documenti – l'HTML – struttura le informazioni in *tag* (i metadati) e *testo* (le informazioni testuali). I tag etichettano il testo, il contenuto, per la sua *rappresentazione visiva* tramite un browser. Gli oggetti rappresentati nel web di documenti sono creati per essere letti, interpretati e, quindi, utilizzati dagli umani, non dalle macchine. Ciò rende assai difficile alle applicazioni software estrarre *frammenti di dati strutturati* dalle pagine HTML per la comprensione e il riutilizzo del dato. Il web di documenti è, dunque, uno spazio informativo globale di testi scritti in HTML (e altri linguaggi di marcatura) collegati da link. Sua caratteristica è essere costituito da oggetti strutturati in una modalità che consente alle macchine l'utilizzo degli elementi informativi ai fini della presentazione formale e logica, ma non permette la *comprensione* del significato degli elementi. Da qui nasce l'idea del web semantico: aumentare la struttura delle informazioni, tramite un processo di atomizzazione che dal *documento* conduce al *dato*, e introdurre ontologie e vocabolari controllati, che consentono alla macchina di arrivare al *significato* del termine. È proprio l'atomizzazione della struttura dell'informazione che esprime le caratteristiche del web di dati: la pagina HTML evolve a favore di dati singoli, particelle minime autonome, ovvero *atomi* (da qui il concetto di *atomizzazione*), riaggregabili con modalità e finalità differenti, *on the fly*, secondo il bisogno, le quali producono a loro

volta nuove entità. L'attribuzione di significato agli elementi, tramite le ontologie e i vocabolari controllati, spiega il meccanismo di *comprensione*.

1.2.2 Il web di dati

I componenti del web di dati (o web semantico) sono:

- le cose (things): cose del mondo reale, come le persone, i luoghi, gli oggetti e i concetti;
- i collegamenti tipizzati (typed links): collegamenti qualificati; nel web semantico chi pubblica i dati rende *esplicita* la *natura della connessione* con altri dati;
- database (A, B, C, ...): contenitori entro cui sono conservate le informazioni poi passate al web.

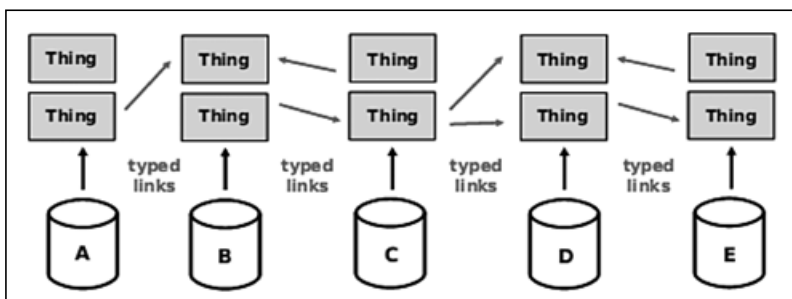


Figura 2 – Rappresentazione del web di dati

Il web semantico è *la naturale evoluzione del web di documenti*. I linked data consentono di creare collegamenti qualificati tra oggetti presenti in diverse fonti di dati, garantendo così legami espliciti tra di essi. Esso è, pertanto, un contenitore di cose (o, meglio, di rappresentazioni di cose) del mondo reale (luoghi, persone, oggetti, ...) e di concetti. I dati afferiscono alla risorsa come attributi e partecipano alla sua natura, ovvero ne sono parte integrante perché la risorsa non sarebbe rappresentabile senza di essi. Nel web semantico cambia la *natura* dei collegamenti, che ora hanno un *significato* specifico, strutturato in una modalità interpretabile e utilizzabile da una macchina. L'aggettivo *semantico* assume la valenza di *comprensibile dalla macchina*: la semantica dei dati consiste nella creazione di informazioni strut-

turate affinché la macchina possa comprenderle e utilizzarle. Il web semantico, secondo Berners-Lee, è “una rete di cose del mondo, descritta dai dati nel web”.⁵ Il concetto è generico, ma contiene alcuni riferimenti importanti:

- la rete (il reticolo);
- le cose (gli oggetti relazionati in modo significativo);
- i dati (non più un record, ma singoli elementi, atomi).

Proprio per le particolarità di questo tipo di web, il concetto di web semantico è strettamente relazionato al concetto di linked data, come metodo e tecnica efficaci per semplificare e omogeneizzare le soluzioni ai problemi d'identità e interoperabilità, con l'individuazione univoca dei dati nel dialogo tra sistemi eterogenei. I linked data sono un complesso di tecniche che, tramite l'uso di vocabolari condivisi, consente la comprensione di un contenuto, pubblicato nel web, da parte di agenti non umani. Ciò testimonia la trasformazione progressiva della rete in un ambiente in cui tutte le risorse disponibili sono associate a informazioni metatestuali che ne specificano il *contesto semantico* in un formato agevole per l'interpretazione automatica e l'utilizzo da parte dei motori di ricerca.

I linked data sono, pertanto, una tecnologia e un insieme di buone pratiche per pubblicare dati nel web in una modalità leggibile, interpretabile e utilizzabile da una macchina, il cui significato sia esplicitamente definito tramite una stringa costituita da parole e marcatori. L'insieme dei dati collegati produce un reticolo: la nuvola (*cloud*).

Il web semantico prevede le seguenti proprietà:

- può contenere qualsiasi tipo di dato;
- i dati possono essere pubblicati da chiunque;
- il dato è autoconsistente, cioè porta in sé la propria descrizione: nel caso in cui un'applicazione incontrasse dati descritti con un vocabolario sconosciuto, può *dereferenziare* l'URI (Uniform Resource Identifier) che identifica ciascun termine del vocabolario e arrivare alla sua definizione: il

⁵ Tim Berners-Lee, Christian Bizer, Tom Heath, *Linked data: the story so far*, <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>.

meccanismo dell'auto-descrizione consente all'applicazione di risolvere l'URI che identifica il termine sconosciuto per recuperarne la definizione;

- le entità collegate da link RDF (Resource Description Framework)⁶ creano un *grafo globale* che si estende e ingloba fonti di dati differenti e consente alle applicazioni la scoperta di nuove fonti informative;
- chi pubblica dati nel web non è obbligato a utilizzare specifici vocabolari per rappresentare i propri dati;
- è aperto: le applicazioni possono esplorare e scoprire nuovi dati seguendo i link del reticolo.

Nell'elenco di proprietà del web semantico sono sottintesi alcuni concetti chiave:

- la correttezza formale e sintattica di una frase pubblicata nel web non implica un'equivalente correttezza semantica: le asserzioni possono essere formalmente corrette e, quindi, valide, ma non necessariamente vere; per esempio, l'asserzione “Immanuel Kant era italiano” è formalmente corretta ma non vera;
- il concetto di dato autoesplicante, che rimanda al concetto di URI dereferenziabile: l'oggetto identificato con l'URI porta con sé una rappresentazione di se stesso e, dunque, diventa comprensibile all'agente che lo incontra nel proprio percorso di esplorazione;
- l'esplorazione del web semantico difficilmente nasce e muore all'interno del medesimo insieme di dati strutturati (dataset): i dati pubblicati in modalità aperta e collegati con relazioni qualificate creano un reticolo che consente di utilizzarli in modo inatteso, non previsto all'inizio dell'esperienza di ricerca, e consentono quei *processi inferenziali* che costituiscono il valore aggiunto del web semantico: a partire da un'asserzione è possibile *dedurre* nuova informazione, senza che essa sia espressamente dichiarata nel dato.

⁶ Linked Data Community, *Linked data. Connect distributed data across the web*, 2012, <<http://linkeddata.org/>>.

1.3 I linked data: definizione in quattro punti

Nel 2006 Berners-Lee propose i *linked data* come metodo elegante ed efficace per semplificare e armonizzare le soluzioni ai problemi d'identità e interoperabilità dei dati. Il metodo fu considerato come “prima espressione concreta del web semantico, utile e ora realizzabile e applicabile a tutte le forme di dati;⁷ esso mirava a creare un *web dei dati* (o delle *entità* o delle *cose*) e si collocava nell'ambito delle tecnologie per il web semantico. La soluzione voleva rispondere ad alcune esigenze del mondo dell'informazione:

- estendere il potenziale del web tradizionale (web di documenti, web ipertestuale), per il quale la semantica del contenuto e i legami tra i documenti sono strutturati dagli umani, ma non possono essere interpretati dalle macchine;
- accedere in modalità integrata ai dati, al di là della loro eterogeneità, quantità, distribuzione e proprietà;
- utilizzare e *riutilizzare* i dati;
- selezionare dati interessanti per la ricerca individuale nel *mare magnum* informativo per rendere possibile la scoperta di dati *pertinenti* all'interno della moltitudine di insiemi di dati disponibili.

“Il web semantico non è solo un modo di mettere i dati nel web. Riguarda il modo di fare collegamenti, cosicché una persona o una macchina possa esplorare il web di dati. Con i linked data, con i dati collegati, se ne abbiamo qualcuno, è possibile trovare altri dati, collegati”.⁸

Berners-Lee definisce quattro regole per la creazione dei dati:

1. usare URI per identificare cose (oggetti);
2. usare HTTP URI in modo che gli oggetti possano essere individuati da persone e da user agent (computer);
3. fornire informazioni sull'oggetto: quando si individua un URI, usando formati standard (RDF, SPARQL);
4. includere link ad altri URI, in modo che possano essere scoperte più cose.

⁷ Linked Data FAQ, Structured Dynamics LLC, USA, <http://structuredynamics.com/linked_data.html>.

⁸ <www.w3.org/DesignIssues/LinkedData.html>.

In un documento del 2006, considerato l'atto fondante del web semantico e della tecnologia dei linked data, Berners-Lee valuta i quattro punti un'*aspettativa di comportamento*, piuttosto che regole di pubblicazione dei dati; sottolinea, infatti, il loro carattere d'*insieme di buone pratiche*.⁹ L'insieme di queste pratiche trasforma il web tradizionale in web semantico; ciò implica:

- l'uso di set universali di simboli per la costruzione di URI: una modalità conosciuta da chiunque si dedichi al web;
- l'uso di HTTP URI: una pratica ampiamente diffusa nel mondo del web; gli URI HTTP sono nomi, cioè identificano la risorsa e non rimandano alla localizzazione della risorsa;¹⁰ sebbene un URI (per esempio, <http://viaf.org/viaf/120688768>, URI con cui l'American Library Association è identificata in VIAF) assomigli sintatticamente a un URL, non intende indicare l'indirizzo di un documento nel web, bensì il suo identificativo;
- l'uso di URI *dereferenziabili*: consentono a un client (computer) di arrivare a una descrizione della risorsa identificata dall'URI; ciascuna risorsa propone una descrizione di se stessa tramite i metadati;
- la creazione di collegamenti con altre risorse esterne: il collegamento tra una risorsa e altre, parte fondante del web tradizionale (ipertestuale), arricchisce la risorsa, aumentando il set informativo delle sue proprietà.

I quattro punti suggeriscono l'utilizzo di identificatori per le cose (le risorse) descritte nel web tramite *metadati*. Gli identifica-

⁹ Il documento, sul sito dell'autore, aggiornato nel 2009, presenta la formulazione: "Status: personal view only. Editing status: imperfect but published", <www.w3.org/DesignIssues/LinkedData.html>.

¹⁰ Vedi: Tom Heath, Christian Bizer, *Linked data: evolving the web into a global data space*. San Rafael, CA: Morgan & Claypool Publisher, 2011, p. 10-11. Per un approfondimento degli HTTP URI usati come nome delle cose, anziché come indirizzi per i documenti nel web, vedi: Norman Walsh, *Names and addresses*, volume 9, issue 73 (25 Jul 2006) (modified 08 Oct 2010), <<http://norman.walsh.name/2006/07/25/namesAndAddresses>>; Henry S. Thompson, David Orchard, *Urns, namespaces, and registries*, <<http://www.w3.org/2001/tag/doc/URNsAndRegistries-50-2006-08-17>>.

tori sono indipendenti da una lingua, più precisi del linguaggio naturale e hanno un valore assoluto.

Nel web semantico cambia, dunque, sia la tipologia di oggetti collegati (dati e non documenti), sia la natura dei collegamenti: mentre nel web di documenti i collegamenti sono solo *link*, nel web semantico i collegamenti hanno un loro *significato specifico*. Karen Coyle spiega la differente natura dei collegamenti in *Linked data tools: connecting on the web*,¹¹ dove esprime il concetto con l'esempio della citazione. Una citazione tradizionale consiste in un numero all'interno di un testo a cui corrisponde una nota a pie' di pagina. Il lettore non conosce il motivo della citazione e può ricavare l'informazione solo dal contesto in cui la citazione è inserita. La medesima citazione nel web semantico è, invece, caratterizzata da un significato specifico; al lettore la citazione risulta esplicita poiché la sua motivazione è espressa nel collegamento stesso: "L'autore cita come prova il testo x" o "l'autore è in disaccordo con il testo x" o "l'autore estende il significato del suo testo", etc.

1.4 Il quadro tecnologico generale

Il web semantico è in continuità con il web tradizionale e di esso sfrutta le tecnologie principali (come l'HTTP URI per identificare gli oggetti), esaltando e strutturando altri meccanismi, come i link qualificati. Il web semantico può essere considerato uno strato strutturato di cui il web si arricchisce, consentendo una maggiore integrazione dei processi macchina nell'utilizzo delle risorse. Il web semantico offre:

- un modello dati (RDF) unico ed estremamente elastico, adattabile a contesti differenti: la possibilità che attori diversi utilizzino il modello per confezionare i propri dati consente una naturale integrazione dei sistemi, senza necessità di ponti, bretelle e traduttori che facciano parlare contesti originariamente eterogenei;
- un modello d'accesso ai dati (tramite URI HTTP) riconosciuto e diffuso nel mondo del web: questa caratteristica

¹¹ Karen Coyle, *Linked data tools: connecting on the web*, report ALA, "Library Technology Reports", vol. 48, no. 4 (May/June 2012).

- garantisce l'esplorazione delle risorse tramite meccanismi tradizionali (browser) e meccanismi di crawler e search engine pensati specificatamente per il web semantico, non ponendo barriere, dunque, alle tecnologie tradizionali;
- il modello dell'interlinking (link tra risorse appartenenti a dataset esterni), che, tramite logiche di costruzione e di utilizzo di URI distribuiti, collega ciascuna entità a entità differenti presenti nel web, attraverso diverse tipologie di relazione (equivalente, gerarchica, etc.);¹²
 - il meccanismo di auto-consistenza dei dati che, tramite vocabolari condivisi, diventano auto esplicanti e comprensibili ad agenti elettronici.

Ciò sembra essere sufficiente a motivare il diffuso interesse delle numerose comunità produttrici o fruitrici dell'informazione verso il web semantico.

1.5 Dal WWW al GGG

Nel novembre 2007 Berners-Lee pubblica un post sul suo blog¹³ in cui parla di *Giant Global Graph* (GGG);¹⁴ si tratta di un nuovo concetto che traccia l'evoluzione della rete “grande quanto il mondo” (il WWW) verso un concetto di “grafo gigante e globale”, ovvero di uno spazio virtuale in cui gli utenti gestiscono il loro “mondo” di relazioni con amici, colleghi e conoscenti. Attualmente, gran parte dell'attività di socializzazione risulta frammentata e ridondante tra più social network (per esempio, Facebook, LinkedIn, Twitter, Jaiku, Live Journal), in cui gli identificativi, le informazioni e i dati prodotti dall'attività di socializzazione (tagging, commenti, punteggi e valutazioni) sono ripetuti.¹⁵ Da questa considerazione nasce la necessità che

¹² Il meccanismo di interlinking, che consente l'integrazione e l'arricchimento delle risorse nel web, è spiegato nel paragrafo 2.6.1.

¹³ <<http://dig.csail.mit.edu/breadcrumbs/node/215>>.

¹⁴ <http://en.wikipedia.org/wiki/Giant_Global_Graph>.

¹⁵ Commenti a post pubblicati sui blog possono essere pubblicati dal medesimo utente su altri blog con l'attività di copia e incolla, moltiplicando e distribuendo in modo esponenziale la medesima informazione sulla rete.

il social web (o web 2.0) incontri le tecnologie del web semantico per ricondurre la frammentarietà a unitarietà. Tim Berners-Lee è diffidente nei confronti degli identificativi numerici degli oggetti e preferisce le sigle alfabetiche. Da qui l'impiego, per esempio, di acronimi quali WWW e GGG rispetto all'uso di identificatori numerici quali web 2.0 e web 3.0.

Se i servizi messi a disposizione dal web sociale adottassero la filosofia del web semantico per la pubblicazione dei dati – uso di formati comuni e di identificatori univoci – lo scambio delle informazioni sarebbe più facile e i dati sarebbero meno frammentati e meno duplicati. Nella terminologia di Berners-Lee, dunque, se il WWW può essere descritto come lo spazio in cui si trova un contenuto con i relativi puntatori (hyperlink non qualificati), il GGG può essere immaginato come lo spazio in cui a tutto ciò si aggiungono le relazioni (i link qualificati) e le descrizioni (riferimento agli URI dereferenziabili): il web semantico.¹⁶

¹⁶ Mauro Guerrini, *Classificazioni del sapere: web semantico, linked data e ontologie. Il ruolo rinnovato delle biblioteche nella trasmissione della conoscenza registrata*, in *Noetica vs informatica. Le nuove strutture della comunicazione scientifica. Atti del convegno internazionale, Roma, 19-20 novembre 2013*, a cura di Fiammetta Sabba. Firenze: Olschki, 2015, p. 145-155.

2. STANDARD, LINGUAGGI E TECNOLOGIE DEI LINKED DATA

Il W3C si occupa di costruire o suggerire uno *stack tecnologico* per la costruzione del web semantico, cioè un insieme di linguaggi, tecnologie, standard e protocolli utili o necessari a chiunque voglia pubblicare i propri dati nel web, costruire vocabolari, scrivere le regole di gestione dei dati secondo i principi dei linked data.

2.1 Identificatori univoci per le risorse

Linked data e web semantico, come si è detto, sono concetti connessi: il primo è una *tecnologia* attualmente adoperata per la realizzazione del secondo. L'identificazione degli oggetti è l'attività qualificante del web semantico. Le biblioteche e gli altri istituti della memoria registrata hanno sempre mostrato attenzione verso la pratica dell'identificazione degli oggetti, creando meccanismi e strumenti d'indicizzazione e controllo, per evitare e risolvere, per esempio, omonimie e sinonimie: la creazione di accessi univoci alle informazioni presenti nel catalogo, la formulazione di voci d'autorità per entità quali nomi, titoli, voci di soggetto, la registrazione di identificatori univoci quali ISBN/ISSN. È questa sensibilità e consuetudine all'identificazione delle risorse che indirizza il web semantico al mondo delle biblioteche, degli archivi e dei musei.

Analizziamo la natura degli identificatori a partire da dati catalografici, con l'esempio di un autore trattato in differenti authority file: la medesima entità – Charles Ammi Cutter – è “un'entità letterale” (testuale, scritta in lettere) nelle schede catalografiche; la forma scelta come accesso al record bibliografico compare in forme varianti in cataloghi diversi:

- A) Cutter, Charles A. (Charles Ammi), 1837-1903 (nel catalogo A)
- B) Cutter, Charles Ammi, 1837-1903 (nel catalogo B)
- C) Cutter, Charles Ammi (nel catalogo C)
- D) Cutter, C.A. (nel catalogo D)

La medesima entità risulta *differentemente identificata* nel catalogo A, B, C e D; un utente potrebbe non riconoscere in queste stringhe la stessa identità; certamente non la riconosce una macchina. La coincidenza di stringhe letterali (nell'esempio tra la voce B e la voce C) risulta un meccanismo d'identificazione molto fragile: la modifica di una voce, anche solo in pochi caratteri della stringa, produce una rottura della relazione d'identità tra le due forme dell'entità. Se le voci sopra riportate fossero, invece, riconosciute come la medesima entità e fossero collegate a un unico identificativo (numerico o di altro tipo), parte del problema sarebbe risolto:

- A) Cutter, Charles A. (Charles Ammi), 1837-1903 => ID: 3567456 (nel catalogo A)
- B) Cutter, Charles Ammi, 1837-1903 => ID: 3567456 (nel catalogo B)
- C) Cutter, Charles Ammi =>ID: 3567456 (nel catalogo C)
- D) Cutter, C.A. =>ID: 3567456 (nel catalogo D)

LC control no.:	n 50027719
LCCN	
permalink:	http://lcn.loc.gov/n50027719
HEADING:	Cutter, Charles A. (Charles Ammi), 1837-1903
000	00534cz a2200169n 450
001	1401655
005	20020516153557.0
008	800605n acannaab a aaa
010	__ a n 50027719
035	__ a (DLC)n 50027719
040	__ a DLC c DLC d MnHi d DLC
100	1_ a Cutter, Charles A. q (Charles Ammi), d 1837-1903
400	1_ w nna a Cutter, Charles Ammi, d 1837-1903
400	1_ a Cutter, C. A. q (Charles Ammi), d 1837-1903
670	__ a His Rules for a dictionary catalog, 1891: b t.p. (Charles A. Cutter)
952	__ a RETRO
953	__ a xx00 b lk19

Figura 3 – Record nell'autority file della Library of Congress, per l'entità Charles Ammi Cutter

Nome autore	Alighieri, Dante
Forme varianti	· Allighieri, Dante · Dante
Datazione	1265-1321
Nota informativa	Poeta. Il vero nome è Durante degli Alighieri. Nato a Firenze e morto a Ravenna.
Fonti	Censimento delle edizioni italiane del secolo XVI. Roma, ICCU: http://edit16.iccu.sbn.it/ Bibliografia nazionale italiana: nuova serie del bollettino delle pubblicazioni italiane ricevute per diritto di stampa a cura della Biblioteca nazionale centrale di Firenze. A. 1, n. 1 (gennaio 1958)- Firenze, Centro nazionale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, 1958- (CDROM)
Regole di catalogazione	RICA
Identificativo SBN	IT\ICCU\CFIV\008732
Identificativo internazionale	IT\ICCU\0000049672
Data di aggiornamento:	08/11/2007

Figura 4 – Record nell’authority file di SBN (Servizio bibliotecario nazionale) per l’entità Dante Alighieri

L’oggetto identificato (ovvero la medesima *cosa*) ha un identificativo univoco (ID: 3567456); per questo motivo l’entità è inequivocabilmente riconoscibile e, pertanto, in contesti differenti è possibile utilizzare voci diverse: editori, distributori, librai, aggregatori di risorse, produttori di bibliografie online, biblioteche. La stringa testuale dell’entità – il testo visualizzato – può essere modificata nel tempo, oppure può avere varianti senza che ciò spezzi il legame semantico tra le fonti (nel caso sopra ricordato, tra i quattro differenti cataloghi). L’identificativo univoco numerico acquista, tuttavia, significato solo se letto nello specifico contesto di produzione.¹

Gli identificativi delle due entità (LC control n° 50027719, Identificativo SBN IT\ICCU\CFIV\008732 e Identificativo internazionale IT\ICCU\0000049672), estrapolati dal contesto catalografico, non sono *riconoscibili* e perdono di significato. L’iden-

¹ Gli ISBN e ISSN, per esempio, perdono significato in contesti non bibliografici o commerciali.

tificazione univoca delle entità, problema irrisolto nel web di documenti, costituisce la tematica centrale del web di dati; ciò impone la scelta di un metodo o di una tecnica che renda la risorsa univocamente identificabile all'interno del proprio dominio d'interesse e nello spazio del web. Nel web semantico l'identificativo della risorsa dev'essere comprensibile a una macchina: lo standard di registrazione dell'ISBN nel formato MARC, per esempio, non prevede un'etichettatura specifica per alcune informazioni aggiuntive (prezzo, supporto, etc.) all'interno della medesima stringa:

024 \$a 9780060723804 (acid-free paper)

024 \$a 9780060799748 (trade)

024 \$a 0394502884 (Random House) \$12.50

Per una macchina diventa impossibile leggere e interpretare il significato del testo entro parentesi (Random House) od oltre esso \$12.50 in quanto non qualificati.

2.1.1 *Identificare gli oggetti con URI*

Ciascun oggetto (entità, risorsa) nel web semantico è identificato da un URI, che rende la risorsa disponibile come link, utilizzabile e dereferenziabile. Per l'entità "Cutter, Charles Ammi", l'HTTP URI attribuitogli da VIAF (Virtual International Authority File) è <http://viaf.org/viaf/69723405/>, che costituisce un identificativo univoco, inequivocabile e dereferenziabile, che può essere usato in contesti differenti. Un URI dovrebbe essere costruito con criteri di semplicità, stabilità, manutenibilità e persistenza a lungo termine;² esso è un *identificativo* piuttosto che un indirizzo della risorsa nel web. Tim Berners-Lee suggerisce di utilizzare HTTP URI (e non uno schema qualsiasi di URI tra i numerosi esistenti) perché essi:

- offrono un modo semplice per creare nomi univoci validi nel contesto globale, con una modalità decentralizzata, senza necessità, cioè, di un'agenzia centralizzata che assegni gli

² Vedi: *Cool URIs for the semantic web* (W3C IG Note), <<http://www.w3.org/TR/cooluris/>>; Tom Heath, Christian Bizer, *Linked data: evolving the web into a global data space*, cit.; Tim Berners-Lee, *Universal Resource Identifiers -- Axioms of web architecture*, 1996, <<http://www.w3.org/DesignIssues/Axioms.html>>..

- identificatori: ciascun responsabile di dominio può creare URI per il proprio ambito d'interesse;
- servono come identificativi e come punto d'accesso alle informazioni che descrivono l'entità identificata: gli HTTP URI possono essere, dunque, *dereferenziabili*.

2.1.2 URI *dereferenziabili* e *content negotiation*

Il meccanismo della dereferenziazione è semplice: un client HTTP dovrebbe poter cercare un URI e recuperare così la *descrizione* della risorsa identificata dall'URI. Il meccanismo riguarda sia gli URI utilizzati per identificare i documenti HTML nel web tradizionale, sia gli URI utilizzati per identificare le risorse nel web semantico. Il web è uno spazio informativo utilizzato da esseri umani e da macchine. Entrambi dovrebbero essere in grado di recuperare le rappresentazioni di risorse in una forma che risponda alle loro esigenze; per questa ragione le descrizioni delle risorse sono presenti nel web in due diverse modalità; in:

- HTML per l'agente umano;
- RDF per la macchina.

Nel caso di URI che identifichino oggetti del mondo reale è importante non confondere l'*oggetto* con il *documento web* che lo descrive: la *data di creazione* di un oggetto o la data di nascita di una persona sono diverse dalla *data di creazione* del documento che descrive tale persona. Per evitare confusione è necessario utilizzare URI differenti, in modo che l'oggetto e la descrizione siano chiaramente identificati e siano differentemente restituiti in relazione all'agente che attivi la richiesta. La differenziazione della risposta è realizzata tramite un meccanismo HTTP chiamato *content negotiation* (negoziiazione del contenuto): i client HTTP inviano *header* HTTP (intestazioni HTTP) che, insieme alla richiesta, indicano (negoziano) quale tipo di formato preferiscono ricevere in risposta. I server che ricevono la richiesta possono analizzarla e selezionare la risposta adeguata:

- se la richiesta indica che il client preferisce l'HTML, il server risponde inviando un documento HTML – per l'occhio umano;
- se il client preferisce RDF, il server invia un file RDF – per la macchina.

L'esempio dell'entità "Cutter, Charles Ammi" e l'HTTP URI utilizzato da VIAF mostra che il medesimo URI (<http://viaf.org/viaf/69723405/>), tramite il meccanismo del content negotiation, rimanda:

- a) alla pagina HTML di descrizione della risorsa (per l'agente umano): nella richiesta inviata al server è indicata l'esigenza d'uso della risorsa e il server restituisce quanto pertinente (la pagina HTML riportata di seguito).

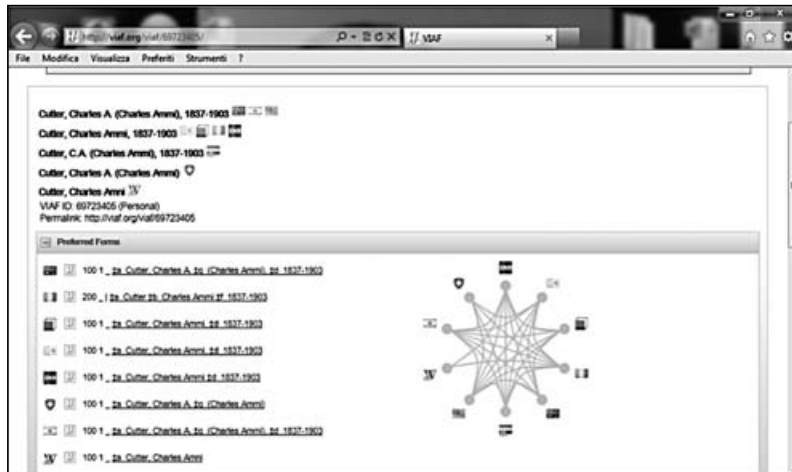


Figura 5 – Pagina HTML di descrizione della risorsa (per l'agente umano)



Figura 6 – Registrazione RDF della risorsa (per la macchina)

- b) alla registrazione RDF della risorsa (per la macchina): nella richiesta inviata al server è indicata l'esigenza d'uso della risorsa e il server restituisce quanto pertinente (la risorsa RDF/XML).

Possiamo ipotizzare per lo meno tre tipologie di URI per ciascuna entità e ciascun URI potrebbe essere costruito in modalità differente:

- l'URI dell'entità in sé, cioè dell'oggetto reale:
 - <http://dbpedia.org/resource/Empoli>
 - <http://id.unifi.it/docenti/Gianni-Rossi>
 - <http://unifi.it/docenti/Gianni-Rossi>
- l'URI che restituisce la descrizione in HTML dell'oggetto reale:
 - <http://dbpedia.org/page/Empoli.html>
 - <http://pages.unifi.it/docenti/Gianni-Rossi.html>
 - <http://unifi.it/docenti/Gianni-Rossi.html>
- l'URI che restituisce la descrizione in RDF/XML dell'oggetto reale:
 - <http://dbpedia.org/data/Empoli.rdf>
 - <http://data.unifi.it/docenti/Gianni-Rossi.rdf>
 - <http://unifi.it/docenti/Gianni-Rossi.rdf>

La logica di costruzione di un URI, pertanto, dev'essere coerente all'interno del medesimo dataset.

2.1.3 *Gli URI: tipologie e proprietà*

Due diverse strategie rendono gli URI dereferenzabili: la strategia URI 303 e la strategia Hash URI; entrambe garantiscono che gli oggetti e le loro descrizioni non siano confusi, e che gli agenti umani e le macchine siano in grado di recuperare la rappresentazione specifica e più appropriata:³

³ Vedi: Tom Heath, Christian Bizer, *Linked data: evolving the web into a global data space*, cit., p. 10-11.

- *URI 303*: usa il meccanismo della ridirezione per restituire i dati RDF come descrizione della risorsa. Esempio di sintassi per questo tipo di URI è: <http://unifi.it/docenti/gianni-rossi>. Nella strategia URI 303, a un client interessato a un oggetto del mondo reale il server risponde con il codice “HTTP 303 *See other*” e con l’URI di un documento web che descrive l’oggetto richiesto. Il messaggio di stato (HTTP 303 *See other*) è chiamato *redirect*: ridirezione verso un altro URI. In una fase successiva, il client dereferenzia il nuovo URI e ottiene un documento che descrive l’oggetto per il quale è originariamente partita la richiesta.
- *Hash URI*: utilizza il carattere hash (#) per introdurre la parte locale o relativa dell’URI, cioè quella che indica la specifica risorsa descritta. Un esempio di hash URI è:
 - <http://unifi.it/docenti#gianni-rossi>
 - <http://unifi.it/docenti#anna-giordani>
- in questo caso il protocollo HTTP risolve la richiesta eliminando la sezione locale dell’URI (#gianni-rossi), restituendo così tutte le risorse appartenenti a uno specifico dominio indicato nell’URI stesso (<http://unifi.it/docenti/>).

Non esistono fattori oggettivi a favore dell’uno o dell’altro tipo di URI, quanto piuttosto considerazioni di opportunità che dipendono dagli specifici contesti di utilizzo. Il limite più rilevante dell’Hash URI è la quantità di dati che l’operazione di dereferenziazione restituisce. Infatti di solito questa tecnica prevede la restituzione di un documento che contiene, al suo interno, una porzione (la parte locale o relativa) di informazione che corrisponde all’URI dereferenziato. È preferibile utilizzare URI 303 nel caso d’interrogazione di dataset corposi e utilizzare Hash URI per l’interrogazione di dataset poco consistenti. L’Hash URI è molto utilizzato per la creazione di vocabolari, i cui singoli termini sono trattati come frammenti significativi.

La logica di costruzione dell’URI è omogenea e diffusa: la stringa è costruita associando a un path iniziale, che identifica il nome del dominio di appartenenza, una parte finale per la se-

zione locale. Esistono alcuni principi generali da rispettare nella costruzione degli identificativi:

- garantire la permanenza nel tempo del nome (cool URIs do not change).⁴ Nell'associare un URI a una risorsa come identificativo univoco non bisogna dimenticare che la risorsa è un oggetto condiviso da una comunità universale, che utilizzerà quell'URI, per esempio, per il meccanismo di interlinking: modificare l'identificativo produce la rottura di quel legame e la negazione del principio d'identificazione;
- non personalizzare per usi locali URI esistenti prodotti da altri soggetti; piuttosto creare nuovi URI, utilizzando logiche di costruzione interne all'ente che li crea, logiche di cui si possa mantenere il controllo nel tempo;
- non utilizzare dati che si riferiscono a dettagli implementativi o identificativi di server: questi dati sono soggetti a cambiamenti nel tempo; per esempio, il cambiamento di un server o del suo hosting produrrebbe modifiche nel puntamento e dunque una situazione di URI non validi. Gli URI dovrebbero essere, quindi, indipendenti o neutrali rispetto a dati tecnici e implementativi;
- utilizzare *chiavi naturali* esistenti come parte locale dell'URI, piuttosto che etichette testuali, le quali, per quanto più leggibili dall'occhio umano, offrono, però, minori garanzie di univocità del dato. Le chiavi naturali più adeguate sono quelle che hanno un significato univoco all'interno di uno specifico dominio: nel caso di risorse bibliografiche, per esempio, costruire l'URI utilizzando numeri di controllo quali ISBN, ISSN, DOI, etc. può essere una buona strategia d'identificazione, anziché utilizzare numerazioni identificative della risorsa all'interno del database in cui si trova; per esempio, la chiave identificativa all'interno di un database. L'utilizzo di identificativi quali ISBN e ISSN per la costruzione dell'URI garantisce una percentuale alta di univocità e una contemporanea facilità di comprensione del dato per gli agenti umani.

⁴ <<http://www.w3.org/Provider/Style/URI>>.

Esistono, tuttavia, casi, piuttosto rari, di ISBN e ISSN non univoci, associati a risorse differenti; alcuni esempi di ISBN che rimandano a titoli diversi:

- 978-84-7788-614-3 è associato a:
 - *Castilla-La Mancha: guía de patrimonio cultural*. Coordinación, Santiago Palomero Plaza, Alfonso Vázquez González. Toledo: Gobierno de Castilla-La Mancha, 2011.
 - *Imaginando el circo: el circo en las colecciones estatales*. Andrés Peláez e Isabel Quintana; edición a cargo del Gobierno de Castilla-La Mancha. Madrid: Gobierno de España, Ministerio de Cultura, Instituto Nacional de las Artes Escénicas y de la Música ; Toledo: Castilla-La Mancha, 2011.
- 978-88-96020-58-6 è associato a:
 - *Il collare di Anàanke*. Carmine Tedeschi. Novi Ligure (AL): Puntoacapo, 2010.
 - *Il metro estensibile*. Ivana Tanzi. Novi Ligure (AL): Puntoacapo, 2010.

2.2 Dai linguaggi di marcatura ai linguaggi ontologici

I linguaggi di marcatura sono utilizzati per arricchire il documento testuale di informazioni aggiuntive, tramite annotazioni, definite *metadati*. Il linguaggio di marcatura utilizzato nel web per rendere questa funzione è l'HTML, un linguaggio di formattazione, che consente di specificare informazioni sulla modalità di presentazione e, dunque, di visualizzazione del documento nel web: parti del documento testuale sono annotate con *tag* (definiti anche *metatag*, riconoscibili perché racchiusi entro parentesi uncinate), i quali garantiscono ai web browser l'identificazione di sezioni di testo e la formattazione secondo le indicazioni fornite. Nell'esempio che segue:

<i>Master biennale di II livello in Archivistica, Biblioteconomia e Codicologia</i>

quanto racchiuso tra i due tag (<i> e </i>) dev'essere visualizzato da un web browser in carattere italico (corsivo):

Master biennale di II livello in Archivistica, Biblioteconomia e Codicologia

Un altro linguaggio che utilizza tag per marcare il documento è l'XML (eXtensible Markup Language), che permette la rappresentazione serializzata di un grafo RDF (nodi e archi); è il tentativo di superare le limitazioni e la scarsa flessibilità di HTML. XML non definisce la visualizzazione del testo, cioè il layout del documento, bensì la sua struttura logica. Esso "costituisce il tentativo di produrre una versione semplificata di Standard Generalized Markup Language (SGML) che consenta di definire in modo semplice nuovi linguaggi di markup da usare in ambito web. Il nome indica, quindi, che si tratta di un linguaggio marcatore (markup language) estensibile (eXtensible) in quanto permette di creare tag personalizzati".⁵ L'XML può essere utilizzato per definire nuovi linguaggi di marcatura; esso struttura il documento tramite elementi racchiusi da tag di apertura e chiusura.

Esempio di documento strutturato in XML:

```
<catalogo>
<cd>
<titolo>Canzoni dell'appartamento</titolo>
  <artista>Morgan</artista>
  <anno>2003</anno>
</cd>

<cd>
<titolo>Non al denaro, non all'amore né al cielo</titolo>
  <artista>Fabrizio de André</artista>
  <anno>1971</anno>
</cd>
</catalogo>
```

⁵ <<http://it.wikipedia.org/wiki/XML>>.

L'XML consente, dunque, di aggiungere struttura e metatag al web, ma è ancora distante dalle esigenze del web semantico, poiché non include informazioni sul *significato* della struttura. Ciascuno può definire tag XML secondo le proprie necessità, senza arricchire i tag di definizioni o significati; le annotazioni XML, da sole, non aggiungono *semantica* al web; affinché un documento XML sia condiviso e *compreso* tra due attori (chi pubblica e chi fruisce) è necessario che la struttura e gli elementi siano conosciuti da entrambi, che abbiano un significato condiviso e dichiarato.

Un passo in avanti verso l'obiettivo del web semantico di esprimere il significato dei documenti è RDF⁶, un modello di dati estremamente flessibile proposto dal W3C per caratterizzare semanticamente le risorse e le relazioni che intercorrono tra di esse. Esso serve per la descrizione dei metadati e consente di costruire asserzioni o dichiarazioni relative ai contenuti del web in termini di proprietà e valori definiti. Il modello di dati si basa su tre concetti:

- risorse identificate da URI;
- proprietà; aspetto specifico, attributo, o relazione usata per descrivere una risorsa;
- asserzioni: una specifica risorsa a cui viene attribuita una proprietà espressa in un valore.

I linked data utilizzano RDF perché *tipizza* e, dunque, *categorizza* meglio di qualsiasi altro linguaggio i legami che intercorrono tra le risorse.

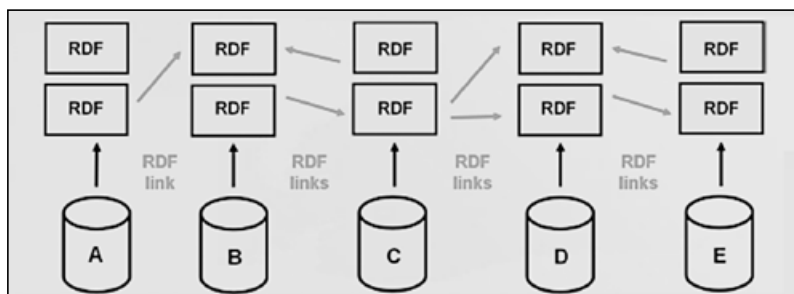


Figura 7 – Mustafa Jarrar, *The data web and linked data*

⁶ RDF è descritto in dettaglio, come parte del W3C RDF Primer, in: <www.w3.org/TR/rdf-primer/>.

2.3 RDF: il modello relazionale dei linked data

RDF è progettato per la rappresentazione integrata di informazioni originate da fonti multiple ed eterogenee, ed è descritto tramite differenti schemi di serializzazione o sintassi (RDF/XML, RDFa, Turtle, N-Triples, etc.). RDF è prima di tutto un modello di dati che, per essere elaborato in modo automatico, dev'essere rappresentato, per esempio, mediante le *triple* (un'asserzione costituita da tre elementi: *soggetto*, *predicato* e *oggetto*, come vedremo poco sotto) o in XML. RDF si propone come una lingua franca, capace di mediare e fare dialogare differenti modelli dati utilizzati nel web.

2.3.1 RDF come modello di dati

Il modello RDF codifica i dati in forma di triple, ovvero di *asserzioni* costituite da:

- *soggetto*: la parte della frase che identifica la cosa, l'entità descritta;
- *predicato*: la proprietà della cosa, dell'entità specificata dalla frase;
- *oggetto*: il valore della proprietà della cosa, dell'entità.

Esempio: Il libro (*soggetto*) parla di (*predicato*) linked data (*oggetto*).

Il predicato è spesso definito *proprietà* per riflettere la sua capacità di descrivere un aspetto del soggetto (quindi della risorsa). Il predicato o proprietà della tripla può essere visto anche come la relazione tra due cose, per esempio, tra un libro specifico e l'editore che lo ha pubblicato.

Soggetto	Predicato	Oggetto
La noia	è pubblicato a	Milano
La noia	ha come editore	Bompiani
Bompiani	è editore di	La noia

Figura 8-a: Esempi di triple RDF

La relazione è *unidirezionale*, e va sempre dal soggetto all'oggetto: questa caratteristica serve anche a identificare bene il soggetto della tripla: nel primo e nel secondo esempio delle triple

riportate *La noia* è il soggetto; nel terzo esempio Bompiani è il soggetto. Per esprimere la relazione bidirezionale tra *La noia* e Bompiani, vanno lette insieme le ultime due triple riportate.

Ogni asserzione è costituita da concetti atomici e significativi: l'unità minima che esprime un concetto in sé compiuto.

Esempi di asserzioni in forma di tripla:

- Alberto Moravia è autore de *La noia*
- Bompiani ha pubblicato *La noia*
- Alberto Moravia è pseudonimo di Alberto Pincherle

Alberto Moravia (*soggetto*) è autore (*predicato*) de *La noia* (*oggetto*).

Il modello RDF pone una serie di restrizioni ai dati che devono essere registrati come triple per garantirne la loro leggibilità da parte dei computer. Il soggetto e il predicato devono essere nella forma di URI; l'oggetto può essere un URI, oppure può essere una stringa leggibile da occhio umano, formata da caratteri, numeri e simboli, chiamata *literal*. Anche l'URI è una stringa di caratteri, numeri, simboli, ma la sua funzione non è di esprimere un significato, né di essere leggibile dall'uomo, bensì di fungere da identificativo unico, universale e leggibile da una macchina. RDF consente inoltre che il soggetto o l'oggetto siano dei *blank node*, anziché URI. Un *blank node* è una stringa identificativa, ma ha valore ed è riconosciuta solo a livello locale. L'uso di un *blank node* invece di un URI è, dunque, sconsigliato, se non in casi strettamente necessari, proprio per la sua natura locale: non è possibile utilizzare un *blank node* per creare una relazione con dataset esterni; la limitazione riduce molto il potenziale di collegamenti esterni con fonti presenti nel web; anche la fusione di dati provenienti da fonti differenti è resa più complicata quando siano utilizzati *blank node* invece di URI come chiave identificativa di una risorsa.

Componente della tripla	Tipologia di dato
Soggetto	URI o <i>blank node</i>
Predicato	URI
Oggetto	URI o <i>blank node</i> o <i>literal</i>

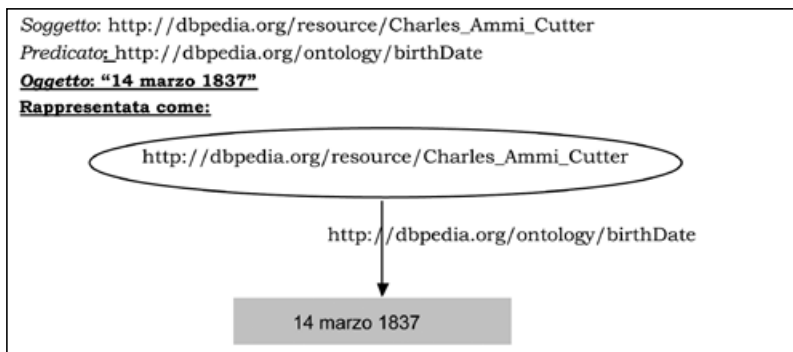
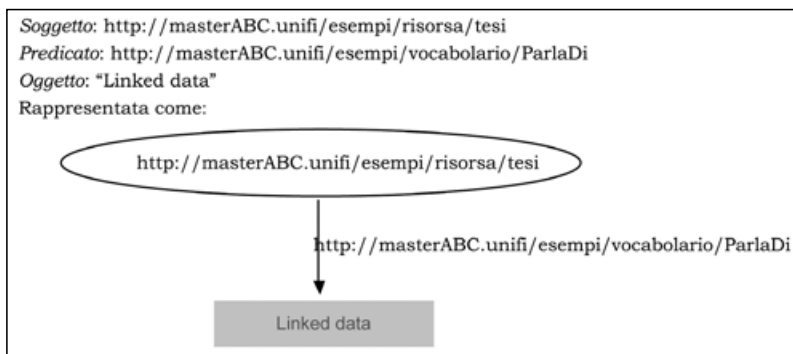
Figura 8-b – Restrizioni sui dati previste da RDF

Sia gli URI, sia i blank node sono dati destinati all'uso da parte dei computer; solo i valori literal sono destinati alla lettura da parte di agenti umani.

2.3.2 Tripla literal

Nei linguaggi di programmazione informatica un *literal* è un simbolo che rappresenta se stesso, una notazione usata per rappresentare una costante, il cui valore può essere un numero (intero o in virgola mobile), un carattere, una sequenza di caratteri (stringa) o un booleano (0 o 1). In RDF i literal sono usati per rappresentare valori come stringhe, numeri e date. La forma più semplice di literal consiste in una stringa Unicode (detta anche *lexical form*). Un literal può essere l'oggetto di una tripla e può essere usato per descrivere le proprietà delle risorse, per esempio, il nome di una persona, il titolo di un'opera, la data di nascita di un autore.

Esempi di tripla con literal:



2.3.3 Link RDF

Un link RDF descrive la relazione tra due risorse tramite *tre* URI, anziché due: ogni elemento della tripla è identificato, infatti, da un URI specifico:

- l'URI *del soggetto* e l'URI *dell'oggetto* verso cui si stabilisce la relazione;
- l'URI *del predicato* che definisce il tipo di relazione tra le due risorse.

L'elemento di distinzione è che anche l'oggetto è costituito da un'entità identificata da un URI e non da un literal. Il predicato è costruito tramite l'utilizzo di vocabolari controllati e condivisi che ne garantiscono la comprensione universale. Tecnicamente, i vocabolari sono collezioni di URI che rappresentano i concetti appartenenti a uno specifico dominio. Qualora l'oggetto della tripla fosse definito come URI rappresenterebbe la chiave del modello a grafo e dell'interlinking, perché consentirebbe di strutturare collegamenti o legami con altri URI; al contrario, una tripla literal interrompe la serie di collegamenti con altre entità, poiché non è identificabile e dereferenzabile dalla macchina.

Esempio di link RDF:

Soggetto: <http://masterABC.unifi/esempi/1234>

Predicato: http://masterABC.unifi/esempi/vocabolario/autore_di

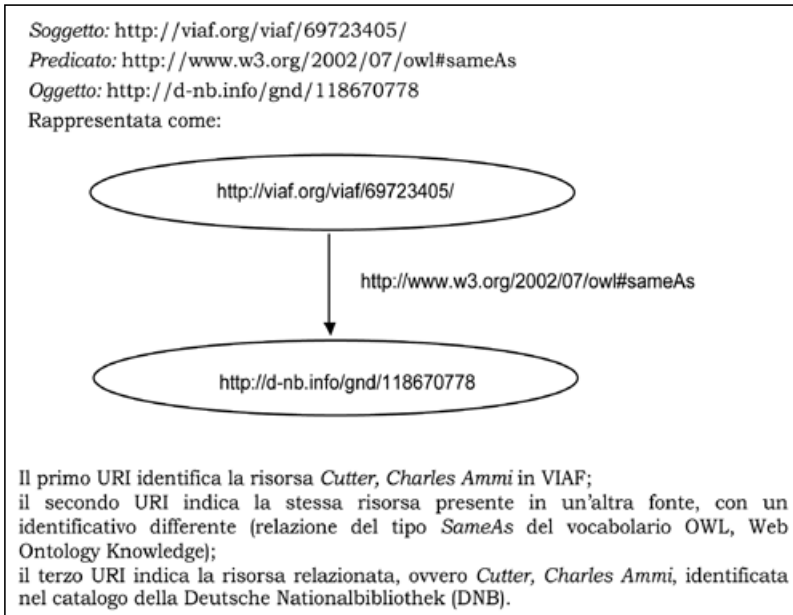
Oggetto: <http://masterABC.unifi/esempi/5678>

il primo URI identifica la risorsa *Alberto Moravia*;

il secondo URI fa riferimento a un vocabolario per esprimere la relazione *autore di*;

il terzo URI identifica la risorsa *La noia*.

Esempio da VIAF:



Più URI sono utilizzati nella costruzione delle triple più l'informazione risulta riutilizzabile. Le triple sono espresse da RDF in forma di *grafi* che rappresentano le risorse, le loro proprietà e i rispettivi valori. In un grafo i *nodi* rappresentano il *soggetto* e l'*oggetto* di una tripla, la linea ad arco che li unisce rappresenta il *predicato*. Il nodo è raffigurato come un ovale se è un URI, come un rettangolo se è un literal: un nodo rappresentato in forma di rettangolo può costituire l'oggetto della tripla, mentre il soggetto e il predicato saranno sempre rappresentati con ovali. Il soggetto e l'oggetto associati a identificatori locali (non URI) sono raffigurati come nodi non etichettati, da cui il nome di *blank node*. Se due nodi di due grafi differenti hanno lo stesso URI possono essere fusi in un unico grafo, senza che la fusione impatti sul significato delle triple in esso contenute: il grafo diventa così un *cluster* (grappolo), cioè un insieme di triple con lo stesso soggetto.

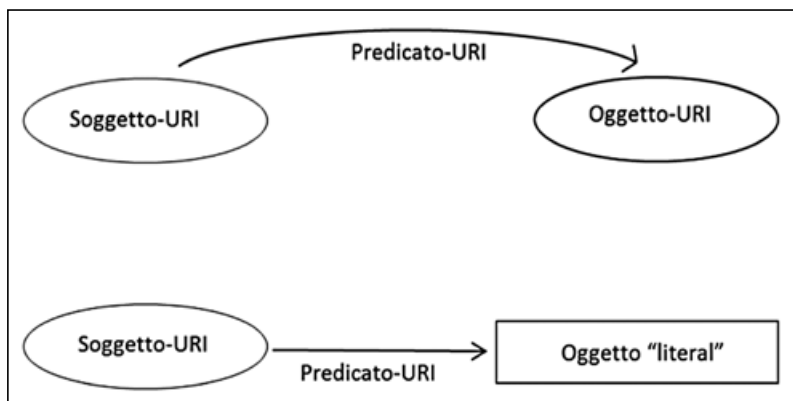


Figura 9 – Grafi RDF costituiti da triple semplici

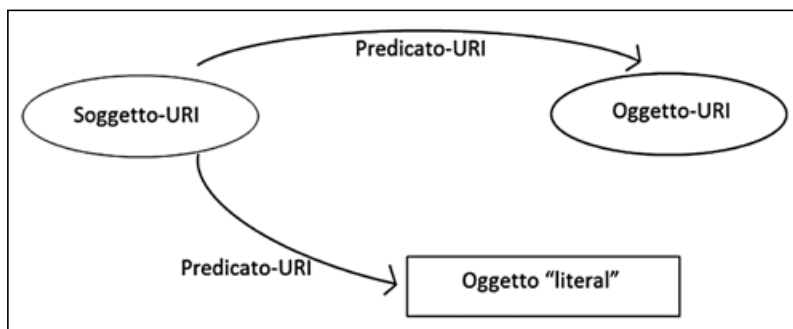


Figura 10 – Cluster: grafo RDF costituito da due triple con lo stesso soggetto

Se l'oggetto di un grafo coincide con il soggetto di un altro grafo, è possibile fondere le due entità in una sola: l'URI comune è l'oggetto di una tripla e nel contempo è il soggetto di una nuova tripla, in una rappresentazione concatenata che prende il nome di *chain* (catena).

Tra le restrizioni e i vincoli posti da RDF nella costruzione di triple, due, riprese da RDF Schema, limitano e controllano la classe di "cose" che possono essere soggetto od oggetto di una tripla basata su una specifica proprietà:

- il *dominio* della proprietà: la classe RDF a cui il soggetto di ogni tripla basata su quella proprietà deve appartenere;
- se una proprietà con etichetta "ha proprietà" ha come do-

- minio la “Classe A”, allora la tripla “soggetto-ha proprietà-oggetto” implica che il soggetto identifica qualcosa, per esempio, un individuo, che è un membro della “Classe A”;
- il *range* della proprietà: la classe RDF a cui l’oggetto di ogni tripla basata su quella proprietà deve appartenere; se la proprietà con etichetta “ha proprietà” ha come range la “Classe B”, la tripla “soggetto-ha proprietà-oggetto” implica che l’oggetto identifica una cosa che è membro della “Classe B”.

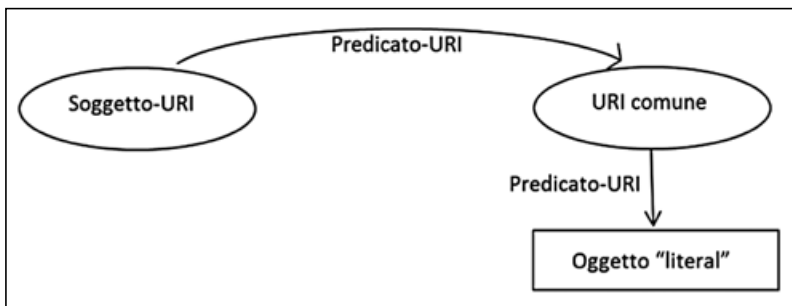


Figura 11 – Grafo RDF a catena, con due triple che hanno in comune soggetto/oggetto

Ciascuna struttura semplice “soggetto-predicato-oggetto” si relaziona e si aggrega con altre strutture semplici, creando così un reticolo più complesso, in cui ciascun soggetto può diventare a sua volta oggetto di altre triple:



Figura 12 – Rappresentazione di una tripla (nodi e archi) in RDF

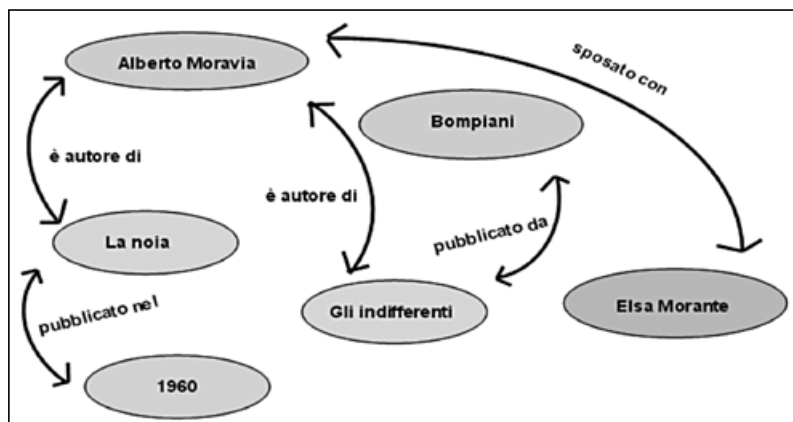


Figura 13 – Rappresentazione di un reticolo di asserzioni o triple

Ogni elemento di ciascuna tripla che costituisce il reticolo può essere a sua volta collegato a elementi di triple di altri reticoli, in una estensione ben più ampia che costituisce una rete nello spazio globale. Questo meccanismo di dati collegati tramite rela-

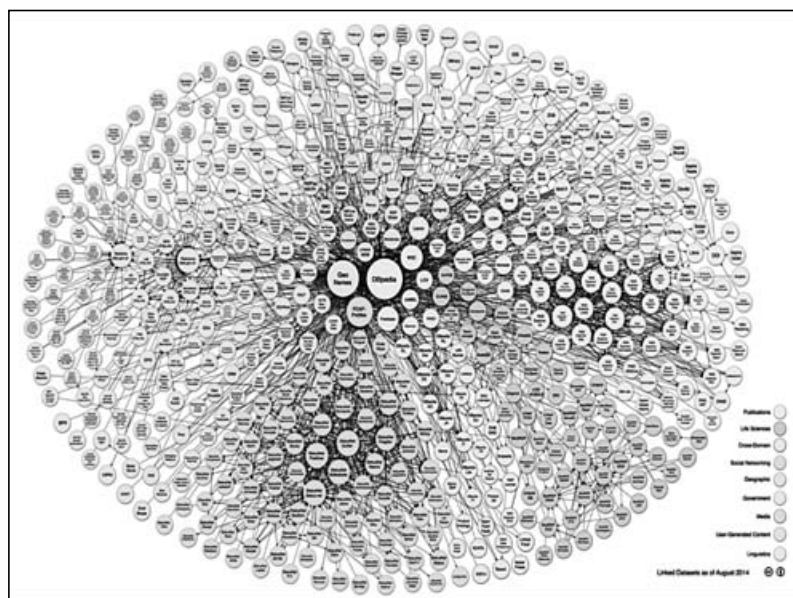


Figura 14 – Diagramma della Linking Open Data cloud (LOD), 30 agosto 2014

zioni significative è l'espressione più efficace dei linked data, in una rappresentazione in forma di grafo che è diventata una sorta di simbolo condiviso del web semantico e della tecnologia dei linked data che lo supporta. Il diagramma della Linking Open Data cloud (LOD)⁷ rappresenta lo spazio web occupato da set di open linked data ad agosto 2014, l'ultimo riprodotto (il primo risaliva al 2007).

2.4 Connettere gli oggetti con relazioni qualificate

Il concetto di *relazione qualificata* è alla base dei linked data: i link connettono le cose ed esprimono *la tipologia di relazione* tra le cose, le entità.

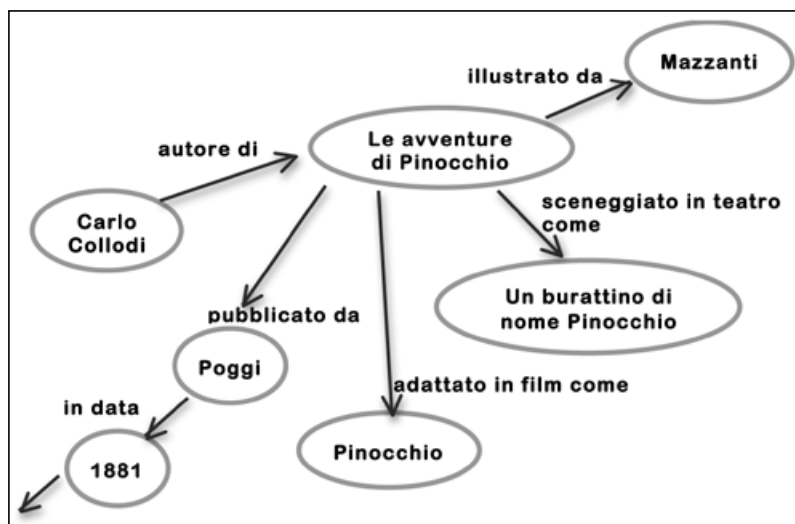


Figura 15 – Esempio di reticolo per *Le avventure di Pinocchio*

Il predicato della tripla esprime questa relazione, che viene resa in un URI per consentire a una macchina d'interpretare la semantica. Il valore del web semantico è espresso nella quarta regola di Berners-Lee: "Includere link ad altri URI, in

⁷ Vedi: Richard Cyganiak and Anja Jentzsch, *Linking Open Data cloud diagram*, <<http://lod-cloud.net/>>.

modo che essi possano scoprire più cose”;⁸ il riferimento è alla possibilità di creare collegamenti tra fonti o dataset differenti tramite *link RDF esterni*, utilizzando l’interlinking, che connette isole di dati nello spazio globale informativo e consente alle applicazioni di scoprire informazioni aggiuntive seguendo i percorsi tracciati dai link.⁹ Il meccanismo del link RDF esterno è semplice: in una tripla RDF il soggetto della tripla è un URI referenziato nel namespace di un dataset, mentre il predicato e l’oggetto della tripla sono URI referenziati nei namespace di altri dataset.

Ciascuna risorsa che compone la tripla può appartenere a un diverso dataset disponibile nel web. Il meccanismo consente di collegare dati presenti in una fonte (link RDF interni) con dati

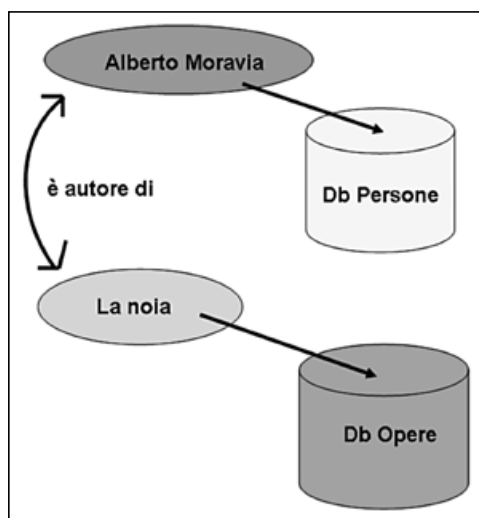


Figura 16 – Esempio di dataset; la tripla può appartenere a un diverso dataset disponibile nel web

presenti in una fonte completamente diversa (link RDF esterni); nell’esempio che segue *Alberto Moravia* è presente nella fonte *Db Persone*, mentre *La noia* è presente nella fonte *Db Opere*; il meccanismo collega i dati della prima con i dati della seconda fonte.

La dereferenziazione di questi URI permette di arrivare alla descrizione della risorsa che può contenere altri link RDF, i quali possono essere derefe-

⁸ Tim Berners-Lee, *Linked data*, <<http://www.w3.org/DesignIssues/LinkedData.html>>.

⁹ Tom Heath, Christian Bizer, *Linked data: evolving the web into a global data space*. California e Colorado, USA: Morgan & Claypool Publishers, 2011, p. 20.

renziati e così via. Il meccanismo consente di passare da una risorsa a un'altra e a un'altra ancora, navigando in un reticolo virtuale. Lo stesso meccanismo di scoperta tramite la navigazione dei link è utilizzato dai motori di ricerca e dai browser.

2.4.1 *Linked data e RDF*

La scelta di strutturare i linked data per mezzo di triple RDF produce notevoli benefici, dovuti soprattutto all'estrema capacità del modello di creare condizioni di condivisione e riutilizzo dei dati nello spazio informativo. L'uso di HTTP URI per identificare in modo univoco le risorse e i vocabolari utilizzati per esprimere le relazioni che intercorrono tra di esse rende RDF un modello universale, utilizzabile in qualsiasi contesto. La capacità di combinare schemi di linguaggio differenti (per esempio, RDFS e OWL) garantisce elasticità nell'utilizzo di RDF per dati molto strutturati e per dati meno ricchi di struttura originaria. In contesti eterogenei, come quelli delle biblioteche (che producono i propri dati per lo più in formati MARC, quindi in modo molto strutturato) e degli archivi (storicamente meno orientati a questa forma di strutturazione) è possibile creare progetti di condivisione superando le barriere dei linguaggi e delle differenti strutture dati.

Le triple RDF devono essere strutturate e mantenute entro *TripleStore* (una sorta di *Semantic Data Base Management System*) e, a loro volta, ospitate da server per interrogare risorse RDF (per esempio, D2RQ Platform for RDF-based access con *mapping per default* in VOID;¹⁰ MySQL;¹¹ PostgreSQL database;¹² Virtuoso¹³), appoggiandosi sul protocollo SPARQL (Protocol and RDF Query Language) endpoint.¹⁴ In prospettiva, con questa architettura, i contenuti digitali modellati nel web come linked data possono essere interrogati in relazione

¹⁰ Vedi: D2RQ - *Accessing Relational Databases as Virtual RDF Graphs*, <<http://d2rq.org/>>.

¹¹ <<http://www.mysql.it/>>.

¹² <<http://www.postgresql.org/>>.

¹³ <<http://semanticweb.org/wiki/Virtuoso>>.

¹⁴ <<http://www.w3.org/wiki/SparqlEndpoints>>.

a “un reticolo sempre più vasto, costituito da ontologie, cioè da *raggruppamenti logicamente definiti*”,¹⁵ contribuendo al miglioramento della “ricerca nel web di altre informazioni affini a quella di partenza: più i dati sono collegati, più sarà possibile il loro utilizzo nell’ottica di arricchimento e deduzione delle informazioni”.¹⁶

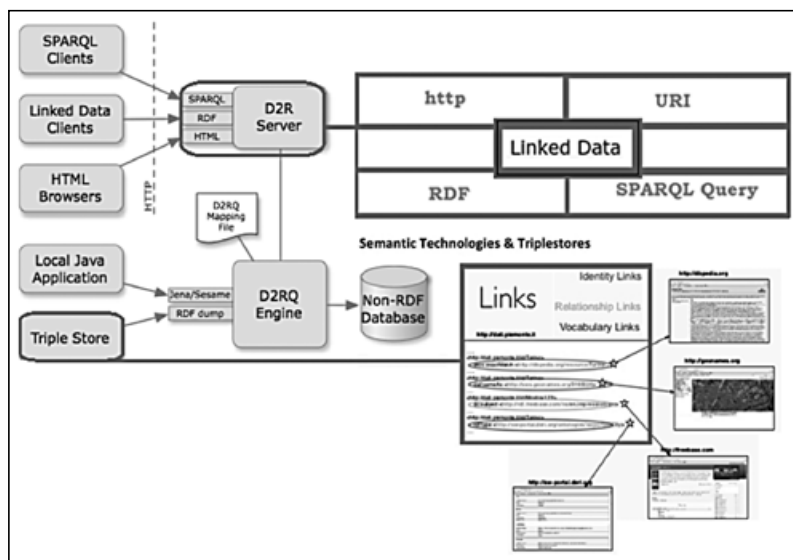


Figura 17 – Esempio di architettura tecnica a supporto della tecnologia linked data

¹⁵ Mauro Guerrini, *Introduzione al seminario “Global interoperability and linked data in libraries”*, “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013), special issue: *Global Interoperability and linked data in libraries*, <<http://leo.cineca.it/index.php/jlis/article/view/6307>>.

¹⁶ Mauro Guerrini, Tiziana Possemato, *Linked data: un nuovo alfabeto del web semantico*, “Biblioteche oggi”, vol. 30, n. 3 (aprile 2012), <www.bibliotecheoggi.it/content/201200300701.pdf>. Riprodotto, con lievi modifiche, in: “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cineca.it/index.php/jlis/article/view/6305>>.

2.4.2 Il web semantico come database globale

La definizione di Berners-Lee del web semantico come un database globale introduce all'analogia con i database relazionali,¹⁷ condivisa dal W3C: "Il modello dati del web semantico è direttamente connesso al modello dei database relazionali. Un database relazionato è costituito da tabelle, realizzate da righe o record. Ogni record è costituito da una serie di campi. Il record non è altro che il contenuto dei suoi campi, proprio come un nodo RDF non è altro che i suoi collegamenti: i valori delle proprietà".¹⁸

L'analogia, cioè la mappatura, tra campi, è diretta:

- un record (la riga) è un nodo RDF;
- il nome del campo (la colonna) è il tipo di proprietà RDF;
- il campo (la singola cella) è il valore.

Cognome	Nome	Età	Telefono
Rossi	Mario	46	06-1234567
Verdi	Antonia	50	06-345678
Bruni	Giovanni	42	06-237890

Lo schema si può spiegare in parole: alla risorsa "Rossi Mario", appartenente alla classe dei dipendenti, sono attribuite le proprietà cognome, nome, età, telefono, formulate tramite i valori "Rossi", "Mario", "46", "06-1234567".

Il punto di forza principale del web semantico è l'espressione della grande quantità di informazioni del database relazionale

¹⁷ Tim Berners-Lee, *Semantic web road map*, 1998, <www.w3.org/DesignIssues/Semantic.html>.

¹⁸ Vedi la sezione *Relational databases on the semantic web*, <www.w3.org/DesignIssues/RDB-RDF.html>.

formulate in una modalità processabile e comprensibile da una macchina, tramite RDF. L'attenzione è posta sugli aspetti essenziali e specifici dei linked data:

- l'atomizzazione dell'informazione, che costituisce il dato o valore contenuto nella singola cella di una tabella;
- le proprietà dei valori (o predicati), formulate tramite il nome del campo che contiene il dato.

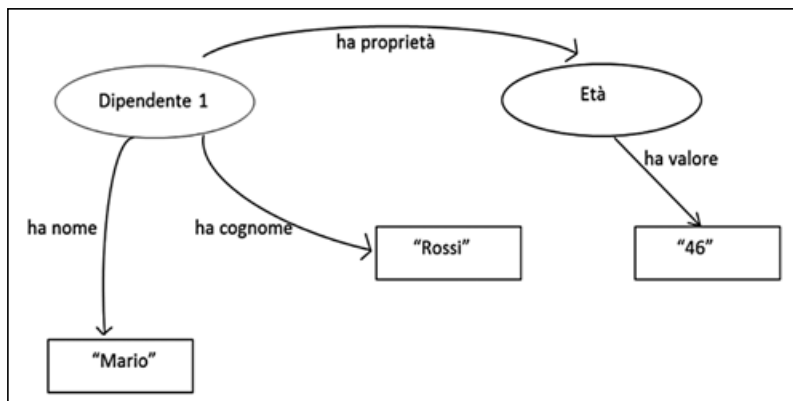


Figura 18 – Rappresentazione in triple di una istanza della tabella dei dipendenti

L'esempio del meccano è calzante: ogni elemento, in sé autonomo, può essere combinato e riusato in una molteplicità di soluzioni.¹⁹

2.4.3 I formati di serializzazione RDF

RDF è un modello di dati per descrivere le risorse in forma di triple costituite da *soggetto-predicato-oggetto*, non è un formato. Per pubblicare le triple nel web è necessario serializzarle tramite una specifica sintassi. Il W3C ha stabilito come standard due formati di serializzazione: RDF/XML e RDFa (Resource Description Framework in Attributes). Esistono molti altri formati utilizzati per la serializzazione di RDF, al di fuori di que-

¹⁹ Vedi: Mauro Guerrini, Tiziana Possemato, *Linked data: un nuovo alfabeto del web semantico*, cit.

sti due riconosciuti dal W3C, tra cui Turtle, N-Triples e RDF/JSON. RDF/XML²⁰ è molto diffuso, ma è complesso da leggere e da scrivere; è, pertanto, poco consigliato quando il workflow di trattamento dei dati richieda un intervento umano rilevante.

Esempio di RDF/XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <rdf:Description rdf:about="http://unifi.it/docenti/gianni-rossi">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person/">
    <foaf:name>Gianni Rossi</foaf:name>
  </rdf:Description>

</rdf:RDF>
```

L'esempio esprime l'esistenza di una risorsa identificata dall'URI <http://unifi.it/docenti/gianni-rossi> del tipo *Persona*, che ha nome Gianni Rossi.

RDFa²¹ è un formato di serializzazione che incorpora le triple RDF in pagine HTML arricchendole con attributi di marcatura. Il risultato è un formato che combina le necessità di pubblicazione dei dati in una modalità leggibile da un agente umano con le necessità di comprensione da una macchina. L'esempio di seguito, tratto dal sito del W3C,²² presenta il caso di una blogger (Alice) che decide di aggiungere informazioni strutturate alle pagine HTML del proprio blog per indicare a una macchina i dati del titolo e della data di pubblicazione di una risorsa. Nella strutturazione dei dati, Alice utilizza il vocabolario Dublin Core.

²⁰ Dettagli sono disponibili su sito del W3C, <www.w3.org/TR/REC-rdf-syntax/>.

²¹ Dettagli sono disponibili su sito del W3C, <www.w3.org/TR/xhtml-rdfa-primer/>.

²² <www.w3.org/TR/xhtml-rdfa-primer/#the-basics-of-rdfa--rdfa-lite>.

Esempio di pagina del blog in HTML:

```
<html>
<head>
...
</head>
<body>
...
<h2>The Trouble with Bob</h2>
<p>Date: 2011-09-10</p>
...
</body>
```

Esempio della stessa pagina del blog in RDFa:

```
<html>
<head>
...
</head>
<body>
...
<h2 property="http://purl.org/dc/terms/title">The Trouble with
Bob</h2>
<p>Date: <span property="http://purl.org/dc/terms/created">2011-
09-10</span></p>
...
</body>
```

Sono riportati in grassetto gli URI di RDFa che strutturano, in una modalità leggibile dalla macchina, il dato sul titolo e il dato sulla data di pubblicazione.

2.5 Tipologie di link

I link RDF possono essere di tipologie differenti, in relazione alla risorsa collegata e alla finalità di collegamento.

2.5.1 *Link di relazione*

I link di relazione collegano risorse diverse in dataset differenti: una persona ad altre persone, a città, a eventi, a istituzioni,

a fatti storici, etc. Questo tipo di link contribuisce ad arricchire il profilo della risorsa di cui si forniscono i dati; è molto utilizzato per descrivere, per esempio, un autore, con i suoi dati biografici, le opere pubblicate, la partecipazione a eventi. I link di relazione creano ponti tra entità che si trovano in dataset differenti, sviluppando una rete virtuale estesa potenzialmente all'intero spazio del web.

Esempio di risorsa utilizzata per costruire un reticolo virtuale:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <rdf:Description rdf:about="http://unifi.it/docenti/gianni-rossi">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person/">
    <foaf:name>GianniRossi</foaf:name> ;
    <foaf:based_near="http://sws.geonames.org/6538558/">
    <foaf:based_near="http://dbpedia.org/resource/Empoli">
    <foaf:knows="http://unifi.it/docenti/giorgio-ruffo">
    <foaf:interest=http://dbpedia.org/page/Functional_Requirements_for_Bibliographic_Records/">
  </rdf:Description>

</rdf:RDF>
```

Figura 19 – Link di relazione e rete virtuale tra dataset differenti

A partire dalla risorsa del tipo *Persona* chiamata *Gianni Rossi* risaliamo, navigando in differenti dataset, alle informazioni sulla città di provenienza (Empoli, di cui abbiamo anche i dati per la geolocalizzazione), sulle conoscenze (Giorgio Ruffo, che è una risorsa contenuta nel medesimo dataset dei docenti dell'Università di Firenze), agli interessi (il rinvio è alla risorsa FRBR nel dataset Dbpedia). A partire da una risorsa (*Gianni Rossi*, che è una persona registrata nel database dei docenti dell'Università di Firenze) riusciamo a navigare e a risalire a informazioni sparse nello spazio globale dei linked data.

2.5.2 *Link d'identità*

I link d'identità collegano la medesima risorsa in dataset differenti: si tratta di un collegamento di tipo *alias*, chiamato anche *URI alias*; essi creano un reticolo che identifica la risorsa seppure descritta in ambiti e modi differenti. Per esempio, *Maria Bianchi* su Facebook è la stessa entità di *Maria Bianchi* su LinkedIn, pur identificata con URI e attributi differenti. VIAF è l'esempio più semplice e comprensibile di questo tipo di URI impiegato nel mondo dell'informazione bibliografica.

Il link d'identità maggiormente utilizzato è il *sameAs*, derivato dall'ontologia OWL, che relaziona, per esempio, la risorsa *Dante Alighieri*, 1265-1321 (identificata sul VIAF con l'URI <http://viaf.org/viaf/97105654>) con altri URI che identificano la stessa risorsa su altri dataset, come data.bnf.fr.²³ La presenza nel web di milioni di predicati formulati tramite il link *owl:sameAs* dimostra la necessità di dichiarare la relazione d'identità per risorse solo apparentemente differenti.

2.5.3 *Link tra vocabolari*

Il link tra vocabolari collega un dato alla definizione di un termine, oppure collega definizioni presenti in vocabolari differenti ma relazionate. Questo tipo di link rende il dato autoesplicante (o autoconsistente) e permette alle applicazioni e ai browser di comprendere la relazione espressa in una tripla tra il soggetto e l'oggetto. Il predicato di una tripla fa necessariamente riferimento a un vocabolario per essere universalmente compreso. Ciò rende possibile l'utilizzo di termini espressi in vocabolari differenti e, quindi, l'integrazione di risorse eterogenee che usino vocabolari differenti. L'integrazione di dati nel web è affrontata in diversi modi:

- utilizzando vocabolari esistenti, senza costruirne di nuovi. Il web è ricco di vocabolari in molti ambiti tematici.²⁴

²³ <http://data.bnf.fr/11898585/dante_alighieri/>.

²⁴ Per un'ampia disamina dei vocabolari, vedi: Library Linked Data Incubator Group, *Datasets, value vocabularies, and metadata element sets*, <<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>>.

- rendendo i dati autoesplicanti, affinché le applicazioni che navigano il web semantico possano risalire alle metainformazioni necessarie per tradurre il dato che proviene da un vocabolario non riconosciuto, in una rappresentazione comprensibile e riutilizzabile. Il riconoscimento si ottiene rendendo gli URI del predicato dereferenziabili, in modo che i client possano risalire alla rappresentazione del termine utilizzato e comprenderla; oppure pubblicando una mappatura tra termini di vocabolari differenti come RDF link; nel caso di termini appartenenti a vocabolari differenti relativi al medesimo concetto è possibile utilizzare alcuni tipi di link RDF che esprimono meglio di altri la relazione di equivalenza, come per esempio l'owl:equivalentClass e l'owl:equivalentProperty.

2.6 Le relazioni nel web semantico

La possibilità di creare collegamenti tra entità in forma di link qualificati è una specificità rilevante del web semantico. È perciò importante creare i collegamenti tra entità per garantire l'integrazione delle risorse nel web e la possibilità che esse siano ricercabili e usabili. Per questo motivo RDA (*Resource Description and Access*),²⁵ primo standard di catalogazione (o, con linguaggio contemporaneo, di metadatazione) che nasce esplicitamente per scoprire le risorse nell'era digitale e che adotta i principi del web semantico, focalizza le linee guida sul concetto di *dato* e sulla creazione di relazioni. Le sezioni 5-10 riguardano, infatti, le relazioni tra le entità FRBR: Opera, Espressione, Manifestazione, Item, Persona, Famiglia ed Ente.²⁶

²⁵ Joint Steering Committee for Development of RDA, *RDA: Resource Description and Access*, 2013, <<http://www.rda-jsc.org/rda.html>>; vedi: Carlo Bianchini, Mauro Guerrini, *Introduzione a RDA*, cit.

²⁶ <<http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>>.

2.6.1 I vantaggi dell'interlinking: arricchimento ed economicità dei dati

Le relazioni sono raramente espresse solo tra risorse all'interno di un medesimo dataset: una caratteristica significativa di RDF e dei linked data è creare ponti con altri insiemi di dati, tramite link qualificati. Il meccanismo dell'interlinking permette di arricchire i propri dati ereditando informazioni da altre fonti. Per esempio, il descrittore *Ermeneutica* del *Nuovo Soggettario*, nella versione RDF/SKOS, è collegato alla medesima voce di *Wikipedia*. Il legame consente l'arricchimento informativo del descrittore con le informazioni di *Wikipedia*.²⁷

Record in HTML:

The screenshot displays the 'Nuovo soggettario THESAURUS' web interface. The header includes the logo of the Biblioteca Nazionale Centrale Firenze and the title 'Nuovo soggettario THESAURUS'. Below the header, there are navigation tabs: 'Criteri', 'Aiuto alla ricerca', 'Sigle e simboli', 'Fonti', 'Novità', and 'Download'. The main content area is titled 'Ermeneutica' and includes a 'Glossario' link. The entry is categorized under 'Azioni Discipline' and includes a 'Nota d'ambito' describing the field of hermeneutics. A table lists related terms: 'UF' (Esergetica), 'TT' (Discipline), 'BT' (Filosofia), and 'RT' (Finitudine Interpretazione). The footer of the entry provides source information (BNI 1956-1985, DeM, PT, LIUC, DDC21, Wikipedia) and classification details (Dewey 121.686, LSH: Hermeneutics).

Biblioteca Nazionale Centrale Firenze
Nuovo soggettario THESAURUS

Criteri Aiuto alla ricerca Sigle e simboli Fonti Novità Download

+ Ricerca per termine del Nuovo soggettario
+ Soggettario(1956)
Lista di aggiornamento (1956-1985)

Sistema Nuovo soggettario
• home page
• manuale applicativo

Ermeneutica [Glossario](#) [RDF/XML](#) [Notizie bibliografiche](#)

Categoria/Faccetta: Azioni Discipline
Nota d'ambito: La scienza e il metodo dell'interpretazione di testi, problemi e quanto altro è oggetto della filosofia, con un'accentuazione della riflessione storico-linguistica, applicata dai filosofi sin dall'antichità, comprende anche le teorie dell'omonima corrente del 20. secolo

UF	Esergetica
TT	Discipline
BT	Filosofia
RT	Finitudine Interpretazione

Fonte: BNI 1956-1985; DeM, PT; LIUC; DDC21; Wikipedia)IT)
Classificazione Dewey (Ed. 22): 121.686
Equiv. LSH: Hermeneutics
Agenzia catalografica/Proponente: BNI
Status del record: Termine strutturato
Identificativo: 2961

²⁷ Vedi: Giovanni Bergamin, Anna Lucarelli, *The Nuovo Soggettario as a service for the linked data world*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cineca.it/index.php/jlis/article/view/5474/7903>>.

Record in RDF/XML:

```
<?xml version="1.0" encoding="UTF-8"?>
- <rdf:RDF xmlns:nsogi="http://prefix.cc/nsogi" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  - <rdf:Description rdf:about="http://purl.org/bnct/tid/2961">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:prefLabel xml:lang="it">Ermeneutica</skos:prefLabel>
    <skos:inScheme rdf:resource="http://purl.org/bnct/tid/ThesCF2"/>
    <skos:inScheme rdf:resource="http://purl.org/bnct/tid/Thes"/>
    <skos:broadner rdf:resource="http://purl.org/bnct/tid/762"/>
    <skos:altLabel>Esegetica</skos:altLabel>
    <skos:related rdf:resource="http://purl.org/bnct/tid/18408"/>
    <skos:related rdf:resource="http://purl.org/bnct/tid/2968"/>
    <skos:closeMatch rdf:resource="http://id.loc.gov/authorities/sh00006024"/>
    <skos:closeMatch rdf:resource="http://id.loc.gov/authorities/sh85060402"/>
    <skos:closeMatch rdf:resource="http://it.dbpedia.org/resource/Ermeneutica"/>
    <skos:closeMatch rdf:resource="http://www.biblio.liuc.it/scripts/thesauro/concept_d.asp?46"/>
    <skos:scopeNote>La scienza e il metodo dell'interpretazione di testi, problemi e quanto altro è oggetto della
      filosofia, con un'accentuazione della riflessione storico-linguistica, applicata dai filosofi sin
      dall'antichità; comprende anche le teorie dell'omonima corrente del 20. secolo</skos:scopeNote>
    <skos:editorialNote>FONTE: BNI 1956-1985; DeM; PT; LIUC; DDC21; Wikipedia(IT)</skos:editorialNote>
    <skos:notation rdf:datatype="http://dewey.info">121.686</skos:notation>
  </rdf:Description>
</rdf:RDF>
```

Una risorsa è, così, identificata e descritta con dati provenienti da dataset differenti: un meccanismo di *arricchimento* delle informazioni combinato a un principio di *economicità* dell'investimento, tramite il riutilizzo di informazioni esistenti. Il meccanismo dell'interlinking aumenta la credibilità e l'autorevolezza del dataset e innesca un circolo virtuoso di condivisione e arricchimento dei dati. Il dettaglio del diagramma dei LOD mostra i numerosissimi link da dataset esterni a fonti quali DBpedia o GeoNames, i dataset più interessanti e utilizzati per l'arricchimento di dati in RDF.

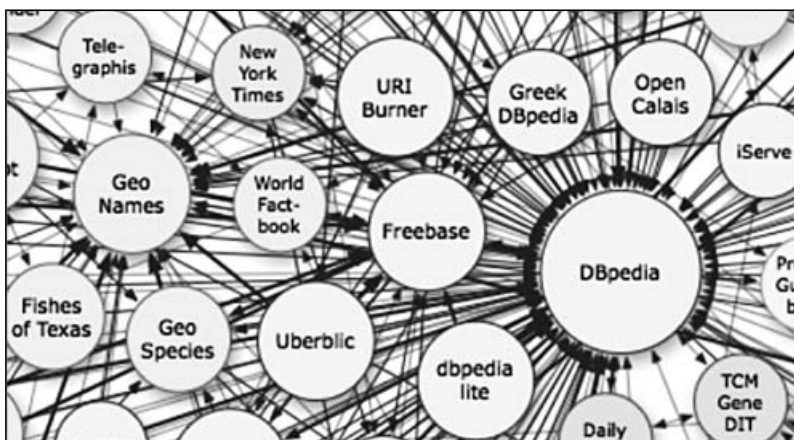


Figura 20 – Dettaglio del Linked Open Data cloud

2.6.2 Proprietà dei dataset

Le raccolte digitali devono possedere una serie di proprietà che ne garantiscano l'affidabilità, riportate nello schema che segue: tra le varie tipologie di raccolte digitali, rientrano i dataset, in quanto metadati, da sottoporre anch'essi a politiche di digital preservation.²⁸

Le principali proprietà a cui deve rispondere un dataset per essere affidabile (trusted) sono:

Proprietà	Descrizione
Autenticità (<i>Authenticity</i>)	Riguarda la certezza che una risorsa digitale appartenga a chi dichiara di esserne creatore e corrisponda a quanto dichiarato. Implica la proprietà di “non ripudio”, cioè la certezza che l'autore non possa negare di essere il creatore dell'oggetto.

²⁸ Lo schema è redatto da Valdo Pasqui. Cfr. Maurizio Lunghi, Chiara Cirinnà, Emanuele Bellini, *Trust and persistence for internet resources*, “Jlis.it.”, vol. 4, no. 1 (gennaio/January 2013); cfr. inoltre: Valdo Pasqui, *La conservazione del digitale: strategie, tecniche, standard ed esperienze per la costruzione di trusted repository di pubblicazioni periodiche*, in *Collezioni digitali di periodici in Italia ed in Europa: standard, applicazioni, valutazioni, prospettive. Atti del convegno, Arco, 15-16 novembre 2007*, a cura di Elena Corradini e Francesca Rocchetti. Trento: Giunta della Provincia autonoma di Trento, 2009, p. 41-56. Per le proprietà e le definizioni vedi il rapporto pubblicato da RLG-NARA nel 2005 e aggiornato nel febbraio 2007 dal Center for Research Libraries (CRL), *Trustworthy Repositories Audit & Certification (TRAC): criteria and checklist*, revised and expanded version of The Audit Checklist for the Certification of Trusted Digital Repositories, originally developed by RLG-NARA Digital Repository Certification Task Force, OCLC and CLR, Version 1, February 2007, <URL: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>>; su TRUST vedi: *TRAC is the basis of the Trusted Digital Repository (TDR) document that was accepted as ISO 16363 in 2012*; e inoltre: Cultural Heritage online - Trusted Digital Repositories & Trusted Professionals. *Rischiamo di perdere la nostra 'memoria digitale'*. Conferenza internazionale promossa dalla Fondazione Rinascimento Digitale dell'Ente Cassa di risparmio di Firenze, <<http://www.apretoscana.org/portal/?q=node/5053>>.

Proprietà	Descrizione
Integrità (<i>Integrity</i>) e controllo dei processi	Concerne l'inalterabilità e la stabilità degli oggetti digitali. Garantire l'integrità dei dati significa prevenire e controllare le perdite e i cambiamenti dovuti a cause accidentali o dolose. Solo controllando i processi lungo il ciclo di vita dell'oggetto digitale si può assicurare la sua integrità e tracciare i suoi cambiamenti o le eventuali minacce ricevute.
Affidabilità (<i>Reliability</i>) e Disponibilità (<i>Availability</i>)	I due concetti riguardano la consistenza e la completezza dei dati. Il primo si riferisce al comportamento dei componenti hardware e software del sistema di conservazione, il quale dev'essere conforme alle loro specifiche, garantendo una percentuale di errori e anomalie minimale. Il secondo riguarda il regolare funzionamento dei sistemi e delle applicazioni, le quali devono produrre costantemente i medesimi risultati per assicurare la corretta fruizione degli oggetti digitali archiviati e dei dati a essi associati.
Accesso, sicurezza e utilizzo continuato (<i>Ongoing usability</i>)	I concetti si riferiscono all'esigenza di garantire l'uso delle risorse digitali nel tempo, assicurandone l'accesso perpetuo, la sicurezza, l'usabilità e la citabilità nel lungo termine. Rientra in questo ambito l'accesso <i>post cancellation</i> , cioè dopo la scadenza della licenza d'uso alle risorse elettroniche acquisite a pagamento e ai dataset.

2.7 Linked data: dataset e namespace

Nell'immagine della sezione di LOD sono rappresentati i dataset RDF Dbpedia, GeoNames, GeoSpecies, Feebase. Alla pagina www.w3.org/wiki/DataSetRDFDumps è registrato un elenco di dataset disponibili come dump RDF, tra cui:

- *Data-govwiki*: dataset RDF di dati convertiti da database pubblicati su <http://data.gov>;
- Dbpedia: dataset RDF di dati estratti da *Wikipedia*;
- LinkedMDB: dataset RDF di dati relativi a film;

- *Nuovo Soggettario*:²⁹ il *Soggettario* aggiornato e strutturato in forma di thesaurus e pubblicato in formato SKOS, realizzato dalla Biblioteca nazionale centrale di Firenze, pubblicato nel 2009, collegato a una serie di altri dataset esterni, tra cui DBpedia e Library of Congress Subject Headings (LCSH);³⁰
- RAMEAU subject headings (*Répertoire d'autorité-matière encyclopédique et alphabétique unifié*)³¹: rappresentazione in SKOS del RAMEAU book indexing vocabulary, creato nel 2008 dalla Bibliothèque nationale de France (BnF); è integrato, tramite meccanismo di interlinking, con il LCSH e con i termini provenienti dalla Deutsche Nationalbibliothek;
- *Deutsche Nationalbibliothek Subject Heading*,³² creato nel 2010 dalla Deutsche Nationalbibliothek; include collegamenti con DBpedia, *Wikipedia* e VIAF.

Le dimensioni di un dataset sono impressionanti: DBpedia contiene, al gennaio 2015, circa 2.7 milioni di concetti descritti da 247 milioni di triple e include abstract in 14 lingue. Si intuisce così la gestione complessa dei dataset, come l'immagazzinamento delle triple, che obbliga chiunque decida di elaborare un progetto di strutturazione di linked data a valutare attentamente tutti gli aspetti tecnologici legati all'hardware e al software.

2.7.1 Creare link RDF

Il processo di creazione di link è a doppio binario, una sorta di attività reciproca che innesca un circolo virtuoso di collegamenti dal dataset di partenza a dataset esterni e viceversa: il processo non è mai univoco o unidirezionale, ma è supportato da una logica di collaborazione reciproca per creare più connessioni possibili tra dataset: più un dataset è collegato, più acquista valore in termini di credibilità e di manutenibilità. Dbpedia, per esempio, è il nodo centrale della nuvola, a cui puntano tutti i dataset che ne fanno parte. La pagina <http://wiki.dbpedia.org/Downloads351#h120-1>

²⁹ *Nuovo Soggettario*, <<http://thes.bncf.firenze.sbn.it>>.

³⁰ <<http://id.loc.gov>>.

³¹ <www.cs.vu.nl/STITCH/rameau>.

³² <<http://wiki.d-nb.de/display/LSD>>.

elenca e consente lo scarico dei dataset integrati in Dbpedia, a disposizione del web e, dunque, usabili da altri editori di dati. L'interlinking è necessario per porre i propri dati *nel web* e non solo *sul web*: un dataset non è mai esaustivo, se pubblicato nel web senza connessioni con altri dataset; la ricchezza di informazioni che si eredita creando legami con altre fonti di dati è impensabile in un contesto di isolamento. Nel sistema a 5 stelle proposto da Berners-Lee per esprimere la qualità del dato pubblicato nel web, il grado di qualità più elevato è dichiarato dalla quinta stella: "Collegare i vostri dati ai dati di altre persone per fornire contesto".³³ L'operazione dell'interlinking, dunque, aggiunge il valore massimo ai dati pubblicati nel web; è per questo motivo che molti creatori di dati adottano la strategia di proporre i propri dati in triple RDF ad altri creatori affinché li integrino in ulteriori dataset.

Un elenco completo dei dataset disponibili per questo tipo di operazione di interlinking è pubblicato in un *data hub*, che mette a disposizione i dati e servizi finalizzati alla condivisione di linked data nel web. La scelta di quali dataset utilizzare per creare link è guidata da varie considerazioni:

- l'affinità del dataset con i propri dati e, dunque, il potenziale di arricchimento che l'integrazione può offrire;
- la stabilità e la manutenibilità del dataset, in considerazione dell'ente che lo pubblica;
- la stabilità e la manutenibilità degli URI che nel dataset identificano le risorse;
- il numero di collegamenti esistenti per il dataset.

Dopo aver selezionato il dataset di proprio interesse, per completare il flusso di creazione di interlinking, è necessario definire i predicati da utilizzare nella struttura del link. Il criterio principale è l'analisi della natura e della tipologia del proprio dato e del dato da collegare; per esempio, i predicati più utilizzati nel dominio delle persone derivano da FOAF; quelli nel dominio dei concetti possono derivare da SKOS.

Nella scelta di un predicato occorre rispettare poche norme basilari:

- il grado di diffusione e utilizzo di un predicato in altri dataset;

³³ <www.w3.org/DesignIssues/LinkedData.html>.

- la qualità del vocabolario, intendendo per qualità il livello di aggiornamento e di manutenzione che il vocabolario garantisce e la caratteristica degli URI, realmente e inequivocabilmente dereferenzabili;
- la garanzia offerta dall'ente che gestisce il vocabolario.

L'interlinking può essere realizzato manualmente o con processi automatici.

Creazione manuale di link RDF

La creazione manuale dei link RDF è consigliata nel caso di dataset poco corposi; il dataset target viene esplorato tramite un'interfaccia di ricerca (che può essere uno SPARQL endpoint, una pagina HTML o un linked data browser) per individuare l'URI d'interesse, da aggiungere al proprio dataset in forma di nuova tripla RDF. La procedura può essere supportata da servizi disponibili su web che offrono indici di URI utilizzabili.³⁴

Auto-generazione di link RDF

Nel caso di dataset corposi è impossibile o sconsigliabile provvedere manualmente alla ricerca e selezione di URI d'interesse; è necessario praticare una strada che consenta la creazione automatica o semiautomatica dei link a risorse esterne. Il meccanismo automatico deve sfruttare l'*intelligenza* delle macchine per individuare nei dataset preventivamente selezionati come interessanti gli identificativi della risorsa da collegare, tramite processi di confronto (*matching*) basati su algoritmi euristici. La tematica dell'individuazione, identificazione e collegamento di medesimi oggetti è ampiamente conosciuta nel mondo dei database, dei motori di ricerca e della costruzione delle strutture ontologiche, e richiama un meccanismo logico definito *risoluzione dell'identità o deduplicazione*. Gli algoritmi di matching per l'individuazione di identificativi degli oggetti si basano su due logiche:

- *una logica semplice*, quale l'uso di ISBN e ISSN per i libri e i seriali: un dataset può esporre questo tipo di identificatore nell'URI o utilizzarlo come proprietà della risorsa; in

³⁴ Tra i più famosi, vedi: Sindice, <<http://sindice.com>> e Falcons, <<http://iws.seu.edu.cn/services/falcons/objectsearch>>.

entrambi i casi l'ISBN o l'ISSN può essere utilizzato per collegare le risorse);

- *una logica complessa* che, in assenza di identificatori univoci, utilizzi un set di parametri (proprietà) per individuare la risorsa d'interesse. Per esempio: se si vuole individuare sul dataset del VIAF la risorsa *Flacco* autore delle *Satire*, si compara con un algoritmo un set predefinito di attributi, che probabilmente sarà esteso alla formulazione *Aulo Persio Flacco* seguito dal titolo *Satirarum liber*. In questi algoritmi, a ogni attributo o gruppo di attributi è assegnato un *peso* per definire una rilevanza delle risorse individuate.

Esempio:

La risorsa *Aulo Persio Flacco*, associata al titolo dell'opera *Satirarum liber*

è più rilevante

della risorsa *Aulo Persio Flacco* associata al secolo di appartenenza, priva del titolo dell'opera *Satirarum liber*.

Gli algoritmi di disambiguazione delle risorse sono implementati in relazione al tipo di risorsa e alla varietà di possibili attributi. Esistono differenti tool (o applicativi) che supportano la creazione di questi algoritmi e la generazione di link RDF basati su queste regole logiche. Uno dei più famosi è SILK Link Discovery Framework, che mette a disposizione un *linguaggio dichiarativo* per la creazione di algoritmi di comparazione ed è compatibile con tutti i dataset che utilizzino il protocollo SPARQL.

2.7.2 Raccomandazioni LODE-BD

Le Raccomandazioni LODE-BD *How to select appropriate encoding strategies for producing Linked Open Data (LOD)-enabled Bibliographic Data* (versione 2.0 del 2013),³⁵ promosse dall'AIMS

³⁵ Imma Subirats-Coll, Marcia L. Zeng, *LODE-BD Recommendations 2.0 : report on how to select appropriate encoding strategies for producing Linked Open Data (LOD) enabled bibliographical data*, last updated 2013, <<http://eprints.rclis.org/22454/>>. Il paragrafo è redatto da Iryna Solodovnik.

(Agricultural Information Management Standards), forniscono le linee guida per arricchire e consentire in maniera automatica di rendere disponibili come linked open data (LOD) dati bibliografici, affinché diventino parte del web semantico. Un aspetto interessante di LODE-BD è l'approccio *decision trees* (alberi decisionali) proposto per individuare i metadati e i loro valori per passare alla fase della loro codifica.

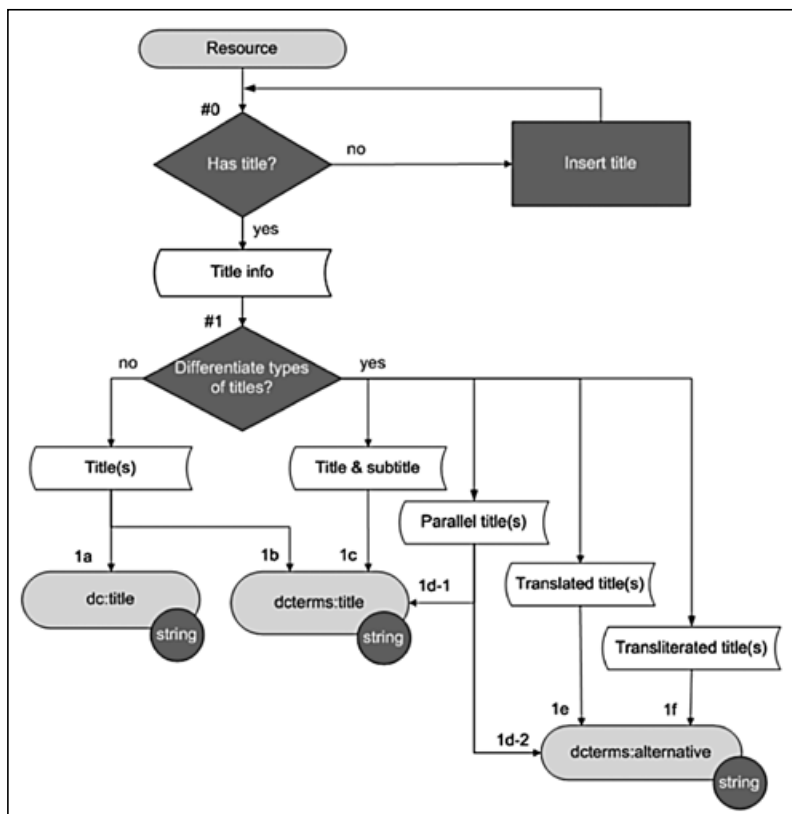


Figura 21 – Esempio di LODE-BD *decision tree* per il metadato *Title*

La figura mostra un esempio di LODE-BD *Decision tree* per il metadato *Title*; esso consiste in un *diagramma di flusso* che guida il processo di scelta di elementi che appartengono a insiemi (schemi) di metadati diversi (per esempio, DC, dcterms) e dei

relativi valori espressi come stringa o come URI. La metodologia si articola in tre macro fasi:

1. individuazione dei tipi di entità e relazioni che devono essere utilizzate per descrivere e accedere alle risorse;
2. definizione delle caratteristiche da soddisfare per la pubblicazione di dati bibliografici come LOD;
3. selezione dei valori adeguati per ogni proprietà.

Ogni *decision tree* è accompagnato da una tabella esplicativa che guida nel dettaglio la scelta dei metadati e dei loro valori, privilegiando l'utilizzo degli schemi di metadati e di domini di valori ampiamente riconosciuti, condivisi e utilizzati a livello globale, per esempio, AGROVOC Thesaurus, VIAF, LCSH e altri schemi modellati come LOD e reperibili da DataHub.³⁶

La modellazione e la successiva codifica dei metadati bibliografici con le raccomandazioni LODE-BD consente di massimizzare l'economicità, il *trust* e il valore semantico delle risorse bibliografiche sul web, rendendole conformi ai formati standard di riferimento per gli *open data* e accessibili in maniera multifocale, così che possono essere incrociati liberamente da diversi sistemi informativi. Grazie ai record d'autorità pubblicati sul web in formato *linked open data*, con cui si raccomanda di codificare i valori di metadati bibliografici, qualsiasi applicazione informatica può accedere a questi dati e riutilizzarli in modo creativo sfruttando le relazioni tra informazioni eterogenee, correlando i percorsi descrittivi delle proprie risorse al grafo LOD dei dati interconnessi mediante i predicati e appoggiati dalle ontologie, contribuendo così in modo automatico alla LOD cloud globale della conoscenza.³⁷

³⁶ *Where to find linked data sets and vocabularies*, in LODE-BD (4. The step forward (With Further Readings), punto 4.3, <<http://aims.fao.org/lode/bd-2/step-forward#where>>.

³⁷ Antonella De Robbio, Silvia Giacomazzi, *Dati aperti con LODE*, "Bibliotime", a. 14, n. 2 (luglio 2011), <<http://www.aib.it/aib/sezioni/emr/bibtime/num-xiv-2/derobbio.htm>>.

3. SCHEMI DI METADATI, VOCABOLARI, ONTOLOGIE

3.1 Vocabolari

RDF fornisce un modello di dati generico e astratto per descrivere le risorse utilizzando triple. Esso non fornisce, tuttavia, i termini di specifici domini, necessari per descrivere le classi delle risorse e le modalità di relazione esistenti tra di esse. La funzione è supportata da tassonomie e vocabolari espressi in linguaggi ontologici. L'uso dei vocabolari nella strutturazione del predicato della tripla attribuisce significato e, dunque, definisce la semantica del dato; è ovviamente necessario che i vocabolari siano condivisi e accessibili alle macchine. Liste di termini controllati e vocabolari sono ampiamente utilizzati in fase di ricerca dell'informazione e di creazione del dato. RDA, per esempio, introduce numerosi vocabolari controllati e condivisi nel web (per la registrazione del contenuto, del formato e del supporto), il cui fine è guidare l'impiego di termini lasciati in passato alla libertà del catalogatore.

Cosa sono questi strumenti e in cosa differiscono?

Vocabolario: lista di termini, prodotta e mantenuta da un ente autorevole; ciascun termine del vocabolario dovrebbe avere un significato univoco, ma non sempre accade; dovrebbero essere rispettate due regole per evitare la:

- *polisemia*: occorre disambiguare lo stesso termine tramite qualificazioni se è usato per indicare concetti differenti in contesti diversi;
- *sinonimia*: occorre indicare un termine come *termine preferito* se più termini sono usati per la medesima entità, gli altri termini sono trattati come *sinonimi*;

Tassonomia: una serie di termini di un vocabolario organizzata in una struttura gerarchica; ciascun termine è relazionato ad altri all'interno della stessa tassonomia tramite relazioni madrefiglia del tipo:

- tutto/parte
- genere/specie
- tipo/istanza

Alcune tassonomie consentono termini *poligerarchici*, appartenenti cioè a differenti termini sovraordinati. In questi casi il termine, pur presente in differenti rami gerarchici, conserva il proprio significato.

Thesaurus: un reticolo di termini di un vocabolario; un thesaurus aggiunge alle relazioni di tipo madre-figlia (viste per le tassonomie), altre tipologie di relazione:

- associative
- equivalenti

Ontologia: esprime differenti concetti, per esempio, glossari, thesauri, tassonomie, schemi, data model, ontologie formali. Un'ontologia formale è un vocabolario controllato espresso tramite uno specifico linguaggio di rappresentazione (un linguaggio logico, ricco e formale, utilizzato per esprimere il significato di ciascun termine). Il linguaggio ha una grammatica che indirizza all'uso di termini del vocabolario per esprimere un significato all'interno di un dominio d'interesse.

3.1.1 Terminologia

La terminologia adottata per esprimere gli schemi di metadati, delle ontologie e dei vocabolari non è univoca, ma dipende dal contesto di utilizzo. Karen Coyle individua quattro concetti, che definisce *elementi*, *ontologie*, *dati* e *vocabolari*, che assumono un significato diverso nel contesto dell'informazione e del web semantico:¹

Ambito tradizionale	Web semantico	Definizione di KarenCoyle
elementi (data elements)	classi proprietà	elementi
schemi di metadati	vocabolari ontologie	ontologie
dati	valori	dati
liste controllate	vocabolari	vocabolari

¹ Karen Coyle, *Linked data tools: connecting on the web*, cit., p. 15.

Dallo schema si deduce che nell'ambito del web semantico l'uso del termine *vocabolario* si sovrappone spesso a quello di *ontologia*, per riferirsi allo schema per la costruzione di vocabolari; con il termine *vocabolario* s'intende, inoltre, l'insieme di lemmi organizzati, tradizionalmente chiamato *lista controllata*. La terminologia non è ancora consolidata, specchio di un contesto in profonda evoluzione concettuale e applicativa. Il W3C Library Linked Data Incubator Group propone sulla pagina *Datasets, value vocabularies, and metadata element sets* del proprio sito una disamina su questi strumenti, anticipata da una definizione di ciascuno di questi termini.²

Set di elementi (element set o metadata element set): definisce le classi e gli attributi (elementi) delle entità. Un element set non definisce entità bibliografiche, ma fornisce gli elementi che devono essere utilizzati per descrivere tali entità.

Termini simili: vocabolario RDF, (RDF) schema, ontologia.

Esempi:

- Dublin Core definisce elementi come il *Creatore* e la *Data*.
- FRBR definisce entità come *Opera* e *Manifestazione* e gli elementi che collegano e descrivono le entità.
- MARC21 definisce elementi (*campi*) per descrivere record bibliografici e d'autorità.
- FOAF definisce elementi che servono per descrivere le persone e che possono essere usati per descrivere gli autori.
- RDA: definisce gli elementi basati sul modello FRBR.

Vocabolario (value vocabulary): un vocabolario definisce concetti (temi, stili artistici, autori) che vengono utilizzati come valori degli elementi nei metadati. In genere un vocabolario non definisce le risorse bibliografiche (per esempio libri) ma i concetti relativi alle risorse bibliografiche (persone, lingue, paesi, etc.). Si tratta di “mattoni” con cui i record possono essere costruiti. Un vocabolario rappresenta, quindi, una “lista chiusa” di valori controllati consentiti per un elemento.

² Vedi: <www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>.

Termini simili: thesaurus, schemi di classificazione, soggettari, tassonomie, vocabolari controllati, authority file, dizionari geografici digitali, schemi concettuali, sistemi di organizzazione della conoscenza.

Esempi:

- LCSH: elenca un insieme di termini utilizzabili per formulare il soggetto di un'opera (per esempio, *Travel*)
- *Nuovo Soggettario*: elenca un insieme di termini utilizzabili per formulare il soggetto di un'opera (per esempio, *Bambini*)
- Art and Architecture Thesaurus (AAT): elenca un insieme di termini specifici relativi a arte e architettura (per esempio, *Impressionist*)
- VIAF: elenca forme di nomi personali (per esempio, *Alessandro Manzoni*)
- GeoNames: elenca luoghi geografici (per esempio, *Roma*).

Dataset: è un insieme di metadati strutturati, descrizioni di cose come i libri di una biblioteca. I record bibliografici sono asserzioni sulle cose, costituite da un elemento (“attributo” o “relazione”) dell'entità e un “valore” per questo elemento. Gli elementi che vengono utilizzati di solito sono selezionati da un insieme di elementi standard, come Dublin Core. I valori per gli elementi sono o presi da vocabolari valore come LCSH, o sono valori a testo libero (ovvero non derivati da vocabolari controllati).

Termini simili: collezioni, set o insieme di metadati.

Esempi:

- un record appartenente a un dataset per un libro può avere un elemento di soggetto tratto da Dublin Core e un valore per il soggetto tratto da LCSH o dal *Nuovo Soggettario*.
- lo stesso dataset può contenere nomi di autori collegati alle loro opere e descritti con elementi tipo il *nome* tratti da FOAF.
- un dataset può essere auto-descrittivo in quanto contiene informazioni su se stesso come entità distinta, per esempio, la data di modifica ed elementi sul curatore, tratti da Dublin Core.

3.2 Selezionare e utilizzare vocabolari per descrivere dati

Nella strutturazione di un dataset, grande importanza ha l'utilizzo di ontologie e vocabolari controllati per la creazione di triple e del predicato in particolare. La logica del web semantico ha definito una linea guida per la selezione, l'impiego e, se necessario, la creazione di nuovi vocabolari e nuove ontologie, edita in un documento più ampio relativo alle buone pratiche per pubblicare i linked data.³ Alcuni criteri generali nell'uso di vocabolari sono:

- scegliere, se possibile, vocabolari esistenti e riconosciuti dalla comunità del web, anziché crearne di nuovi;
- attenersi a regole condivise per garantire la comprensione e il riutilizzo del proprio dataset nel web.

La scelta di uno o più vocabolari è molto delicata, perché i predicati delle triple dipendono dalle definizioni dei termini. Nella scelta del vocabolario occorre tener presente alcune caratteristiche; essi dovrebbero essere:

- *documentati*: dotati di pagine descrittive sulle classi e sulle proprietà utilizzate;
- *auto-descrittivi*: ciascuna proprietà o termine in un vocabolario dovrebbe riportare un'etichetta, una definizione e un commento, in modo da diventare auto-parlante;
- *plurilingue*; tutti gli elementi di un vocabolario, anche quelli rappresentati da codici o identificativi numerici, dovrebbero presentare etichette, definizioni e commenti in più lingue;
- *riusabili*: utilizzati in più dataset per garantire la riusabilità del vocabolario, caratteristica fondamentale nei linked open data;
- *autorevoli*: pubblicati da enti qualificati, che ne garantiscano la qualità del contenuto e l'accessibilità nel tempo;
- *accessibili a lungo termine*: dovrebbero garantire caratteristiche di manutenibilità, persistenza e utilizzo nel tempo;

³ Vedi: *Best practices for publishing linked data:W3C editor's draft 26*, March 2013, <<https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>>.

- *dotati di URL persistenti*: dovrebbero garantire l'accesso continuativo nel tempo al server che ospita il vocabolario, per facilitarne il riutilizzo;
- *aggiornati*: dovrebbero assicurare una politica di versionamento costante; il produttore dovrebbe mantenere aggiornato il vocabolario e assicurare la compatibilità con versioni differenti pubblicate negli anni; i cambiamenti più importanti nei vocabolari andrebbero documentati.

3.3 Linguaggi per le ontologie nel web semantico

In ambito culturale esiste una solida tradizione d'uso di sistemi di classificazione, vocabolari controllati e authority file impiegati soprattutto nella strutturazione e standardizzazione del dato in fase di creazione e di recupero. Le ontologie arricchiscono questa tradizione, favorendo la strutturazione dei dati in modalità utilizzabili dalle macchine, promuovendo così progetti di cooperazione e d'integrazione tra enti di natura e tipologia differenti. I vocabolari servono per classificare i termini utilizzabili in una particolare applicazione, caratterizzare le relazioni, definire vincoli sull'uso di questi termini. Essi possono essere complessi (con migliaia di termini) o semplici (con pochi concetti).

Le ontologie sono utilizzate in sistemi in cui i concetti debbano essere identificati univocamente e in cui agenti software possano riconoscere questi oggetti e realizzare associazioni ed equivalenze. Le ontologie sono, quindi, schemi di attributi e relazioni tra entità; esse consentono di rappresentare le risorse tramite la descrizione delle loro caratteristiche e l'identificazione delle relazioni esistenti tra di esse, e dunque di esprimere la semantica che lega queste entità, per realizzare *categorizzazioni* e processi inferenziali (ragionamenti deduttivi).³¹ In quanto strumenti formalizzati, le ontologie seguono regole espresse nei linguaggi semantici.

⁴ Vedi: Tim Berners-Lee, James Hendler, Ora Lassilla, *The semantic web*, "Scientific American", 17 May 2001, <<http://www.scientificamerican.com/article/the-semantic-web/>>.

I linguaggi semantici più utilizzati nell'ambito del web semantico sono:

- SKOS (Simple Knowledge Organization System)
- OWL (Web Ontology Language)
- RDFS (RDF Vocabulary Description Language), conosciuto anche come RDF Schema.

SKOS è un linguaggio impiegato per esprimere concetti gerarchici di equivalenza e associativi, spesso riferiti a tassonomie, thesauri, soggettari. È utilizzato per rappresentare, per esempio, la relazione tra il concetto di *meccanica* e il concetto più ampio (sovraordinato) di *fisica*.

Esempio di utilizzo dello SKOS nel *Nuovo Soggettario*, per il termine *Autocoscienza*:

```
<?xml version="1.0" encoding="UTF-8"?>
- <rdf:RDF xmlns:nsogi="http://prefix.cc/nsogi" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:dc="http:
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:skos="http://www.w3.org/2004/02/skos"
  - <rdf:Description rdf:about="http://purl.org/bnct/tid/29959">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:prefLabel xml:lang="it">Autocoscienza</skos:prefLabel>
    <skos:inScheme rdf:resource="http://purl.org/bnct/tid/ThesCF8"/>
    <skos:inScheme rdf:resource="http://purl.org/bnct/tid/Thes"/>
    <skos:broader rdf:resource="http://purl.org/bnct/tid/820"/>
    <skos:related rdf:resource="http://purl.org/bnct/tid/11402"/>
    <skos:related rdf:resource="http://purl.org/bnct/tid/762"/>
    <skos:related rdf:resource="http://purl.org/bnct/tid/1982"/>
    <skos:related rdf:resource="http://purl.org/bnct/tid/21561"/>
    <skos:editorialNote>FONTE: Soggettario; VLI; DeM; Zin; Abbagnano; EGF; EPG</skos:editorialNote>
  </rdf:Description>
</rdf:RDF>
```

OWL e RDFS sono linguaggi che descrivono modelli concettuali in termini di classi e proprietà; è possibile utilizzare un vocabolario RDFS per definire il concetto di *Animali* che include la classe dei *Cani* a cui tutti gli esemplari di cane appartengono. È, inoltre, possibile definire la proprietà *Ha colore* che serve per aggiungere attributi informativi se si desidera descrivere il colore di un cane. Sono strumenti che vengono utilizzati soprattutto nei casi in cui si desideri esprimere rapporti di *sussunzione* tra i termini: “Tutti gli atleti *sono anche* persone”. Se utilizzati da opportuni *reasoning engine* (motori che usano regole logiche per realizzare ragionamenti), essi riescono a produrre processi *inferenziali* e, dunque, a *dedurre conseguenze logiche* a partire da asserzioni.

3.3.1 Schemi di dati e ontologie

I dati dei record bibliografici sono solo in parte *di settore*, ovvero specifici: buona parte di essi si sovrappone a elementi utilizzati nel web in contesti differenti. Karen Coyle schematizza vocabolari e ontologie, secondo una classificazione di ambito di studio e di applicazione:

- a) Schemi di metadati per le risorse nel web
 - *DCMI Metadata Terms*:⁵ è lo schema dati del Dublin Core, il cui nucleo è stato elaborato nel 1996; è considerato la madre di tutti i metadati⁶ perché è stato il primo dedicato al mondo dell'informazione bibliografica (editori, provider, biblioteche, etc.); mantiene un ruolo di riferimento nella descrizione delle risorse nel web per la sua essenzialità.
 - *Schema.RDFS.org*:⁷ Bing, Google e Yahoo!, nel giugno 2011, hanno introdotto e condiviso *Schema.org*⁸ quale strumento di costruzione delle pagine web; un insieme di termini (definito anche *microformat*) che i webmaster possono utilizzare per formattare le pagine, aumentando le funzionalità dei motori e la visualizzazione dei risultati della ricerca. *Schema.Org* è stato considerato nelle Raccomandazioni *LODE-BD 2.0*⁹ per provvedere alla mappatura tra i marcatori dello schema, utilizzati da diversi agenti web, con una serie di metadati offerti dalle Raccomandazioni stesse, per creare micro-dati dalle singole fonti in modo da esporle il meglio possibile nel web. La comunità dei linked data ha elaborato successivamente *Schema.RDFS.org*, per incrementare le funzionalità di ricerca e visualizzazione delle pagine nel web semantico. *Schema.RDFS.org* aggiunge la possibilità di creare link e, dunque, di definire relazioni tra risorse nel web, caratteristica qualificante il web semantico.

⁵ <<http://dublincore.org/documents/dcmi-terms>>.

⁶ Vedi: Karen Coyle, *Linked data tools: connecting on the web*, cit., p. 17.

⁷ <<http://schema.rdfs.org>>.

⁸ <<http://schema.org/>>.

⁹ <<http://aims.fao.org/lode/bd>>.

- *RDF Schema*:¹⁰ è stato sviluppato dal RDF Core Working Group come parte del W3C Semantic Web Activity (Activity Statement, Group Charter); è stato utilizzato in progetti qualificati come *data.bnf.fr*.¹¹ Lo schema è strutturato in 6 classi (i gruppi in cui le risorse, o istanze, sono divise) e 7 proprietà (che esprimono la *relazione* tra il soggetto e l'oggetto).

b) Ontologie per la *descrizione bibliografica e archivistica*

- *FRBR (FRBR Core*,¹² *FRBR Extended*,¹³ *FRBRoo*,¹⁴ *FRBRer Model*):¹⁵ il modello concettuale FRBR, emanato dall'IFLA (International Federation of Library Information and Institutions), è stato tra i primi a essere espresso in RDF e ha avuto differenti versioni, alcune ufficialmente riconosciute dall'IFLA (FRBRer) e altre non ufficiali realizzate quando il fenomeno dei linked data è esploso. FRBRer è l'ontologia emanata dall'IFLA FRBR Study Group, inserita nell'Open Metadata Registry (OMR). È costituita da 10 classi (corrispondenti alle 10 entità di FRBR¹⁶) e 206 elementi (che includono gli attributi e le relazioni di FRBR, come definite nella versione del *Report* del 2008). I 206 elementi sono stati discussi nel 2010 e approvati nel 2011. FRBRoo è un'armonizzazione tra FRBRer e CIDOC CRM;¹⁷ si focalizza sulla descrizione di informazioni bibliografiche e museali in progetti d'integrazione dei dati culturali nel web.

¹⁰ <<http://www.w3.org/TR/rdf-schema/>>.

¹¹ Il progetto *data.bnf.fr* è consultabile all'indirizzo <<http://data.bnf.fr>>; è un riferimento essenziale per ogni ipotesi di progetto di linked data, anche per la particolare presentazione dei dati secondo il modello FRBR.

¹² <<http://vocab.org/frbr/core.html>>.

¹³ <<http://vocab.org/frbr/extended.html>>.

¹⁴ <www.cidoc-crm.org/frbr_inro.html>.

¹⁵ <<http://iflstandards.info/ns/fr/frbr/frbrer>>.

¹⁶ Con la pubblicazione di *Functional Requirements for Authority Data (FRAD)*, München: K.G. Saur, 2009, le entità diventano 11, con l'inserimento di *Famiglia*, nel Gruppo 2; la novità non è stata recepita dall'ontologia.

¹⁷ <<http://www.cidoc-crm.org/>>.

- *ISBD elements*:¹⁸ risultato del lavoro congiunto tra l'ISBD Review Group e l'ISBD/XML Study Group, presieduto da Gordon Dunsire, è la traduzione di ISBD (*International Standard Bibliographic Description*) in elementi RDF; riproduce la struttura in 8 aree dello standard.¹⁹ È costituita da 181 classi e proprietà, o più precisamente da 181 *sottoclassi* e proprietà, in quanto ISBD si riferisce all'entità manifestazione di FRBR e definisce, dunque, la sola classe "Risorsa", coincidente con la manifestazione. È pubblicata sull'Open Metadata Registry e utilizzata in progetti internazionali, per esempio, nel data model della British Library, così come nel progetto spagnolo data.bne.es.
- *RDA elements*:²⁰ prodotto dal Dublin Core Metadata Initiative (DCMI) e dallo Joint Steering Committee for Development of RDA (JSC), è stato elaborato tra il 2008 e il 2010; include circa 1300 elementi, raggruppati, per semplicità di consultazione, secondo le entità FRBR; lo schema è pubblicato sull'Open Metadata Registry.
- *Ontologia di EAC-CPF*: l'Archivio centrale dello Stato e l'Istituto beni culturali della regione Emilia-Romagna, con la collaborazione tecnica della società Regesta.exe, hanno avviato un progetto di sperimentazione dei linked open data in ambito archivistico, chiamato Reload, trasformando in ontologia lo standard dedicato all'identificazione dei soggetti produttori EAC-CPF (Encoded Archival Context for Corporate Bodies, Persons and Families);²¹ il 25 ottobre 2011 il Library Linked Data Incubator Group del W3C ha inserito l'ontologia EAC-CPF tra i vocabolari disponibili nel report *Datasets, value vocabularies e metadata element sets*.
- *Ontologia OAD*: è dedicata alla descrizione archivistica e utilizzata nel progetto Reload.

¹⁸ <<http://iflastandards.info/ns/isbd/elements>>.

¹⁹ In ISBD, Edizione consolidata (edizione originale inglese: München: K.G. Saur, 2011; traduzione italiana: Roma: ICCU, 2012), le aree diventano 9, con l'inserimento dell'area 0.

²⁰ <<http://rdvocab.info>>.

²¹ Informazioni più approfondite sono disponibili al link <<http://eac.staatsbibliothek-berlin.de/>>.

c) Ontologie per gli *elementi d'autorità*

- *Functional Requirements for Authority Data (FRAD)*:²² lo schema FRAD, pubblicato sull'Open Metadata Registry, esprime in RDF classi e proprietà del modello dati da cui deriva.
- *Functional Requirements for Subject Authority Data (FRSAD)*:²³ lo schema prevede 2 entità (*Thema* e *Nomen*) e 19 elementi e relazioni (tipo di tema, nota di scopo, lingua, stato, soggetto, etc.); è stato emanato dall'IFLA Working Group on FRSAD e pubblicato sull'Open Metadata Registry.
- *MADS in RDF*:²⁴ l'ontologia è prodotta e mantenuta dalla Library of Congress, per convertire in RDF la struttura complessa del LCSH (Library of Congress Subject Headings). Essa riproduce lo schema di elementi basati sul MADS (Metadata Authority Description Schema).²⁵ L'ontologia serve per trattare voci complesse come quelle del LCSH e accessi Nome/Titolo.

d) Ontologie per la descrizione delle *persone*

- *Friend of a Friend (FOAF)*:²⁶ l'ontologia è stata concepita nel 2000 e pubblicata nel 2008 per codificare informazioni sulle persone nei social network; è diventata il primo strumento di creazione dei metadati di persone nel web semantico; si è diffusa nel mondo delle biblioteche e degli archivi come schema di metadati utilizzato per la codifica di informazioni sulle persone, i luoghi a esse associati, le relazioni con altre persone. Introdotta per i social network presenta lacune che, per il mondo delle biblioteche, la limitano molto; per esempio, consente di esprimere gli attributi del giorno e del mese di nascita di una persona (elementi utilizzati nei social network per condividere compleanni), ma non l'anno di nascita e di morte. Come altri strumenti nati in ambito generico e successivamente estesi a contesti specifici, gli elementi distintivi dovrebbero essere aggiunti dalle comunità

²² <<http://iflastandards.info/ns/fr/frad>>.

²³ <<http://iflastandards.info/ns/fr/frsad>>.

²⁴ <<http://www.loc.gov/standards/mads/rdf>>.

²⁵ Per dettagli sullo schema, vedi: <www.loc.gov/standards/mads/>.

²⁶ <<http://xmlns.com/foaf/spec>>.

specializzate. La comunità bibliotecaria, per esempio, potrebbe integrare le specificazioni circa la forma preferita del nome di una persona o di un ente rispetto a una o più forme varianti, operazione molto diffusa in ambito bibliografico (authority control).

- *BIO*:²⁷ contiene termini utili per aggiungere qualifiche alle persone e informazioni sui contesti sociali in cui sono inserite, nonché relazioni con dati genealogici. È focalizzata sulla registrazione degli eventi significativi della vita di una persona (nascita, percorso personale, educativo e professionale, come diploma conseguito, laurea, matrimonio, divorzio, morte, etc.); è utilizzata con l'ontologia *RELATIONSHIP* ed entrambe completano gli elementi descrittivi di FOAF.
- *RELATIONSHIP*:²⁸ è un vocabolario per la descrizione di relazioni tra persone; aggiunge a *BIO* elementi quali *conoscenze*, relazioni tipo *moglie di*, *compagno di*, etc.

Ontologie per la descrizione di *luoghi geografici*

- *GeoNames Ontology*:²⁹ è tra le ontologie più usate nel web; permette di aggiungere informazioni semantiche geospaziali. Gli oltre 8,3 milioni di toponimi presenti hanno un URL univoco e un web service RDF per essere richiamati da applicativi esterni. Altri web service descrivono le relazioni tra toponimi. È uno dei dataset più collegati da dataset esterni. I luoghi sono categorizzati in città, monti, laghi, continenti. L'ontologia è usata, tra l'altro, dagli OPAC e dai discovery tool per mostrare la localizzazione geografica della biblioteca.
- *FAO Geopolitical Ontology*:³⁰ la FAO (Food and Agriculture Organization of the United Nations) è stata tra le prime organizzazioni a condividere dati e servizi nel web semantico. L'ontologia pone particolare attenzione su aspetti geografici, demografici, governativi ed economici. Le risorse sono categorizzate in: entità auto-governanti, non auto-gover-

²⁷ <<http://vocab.org/bio/0.1/.html>>.

²⁸ <<http://vocab.org/relationship/.html>>.

²⁹ <<http://www.geonames.org/ontology/documentation.html>>.

³⁰ <www.fao.org/countryprofiles/geoinfo/geopolitical/resource>.

nanti, altre organizzazioni, regioni geografiche, gruppi speciali, regioni economiche. L'ontologia è molto utile per le biblioteche che trattano discipline politiche, economiche e demografiche.

f) Ontologie per la descrizione dei *diritti* (di uso, di distribuzione, etc.)

- *Creative Commons Rights Expression Language* (CC REL):³¹ descrive la titolarità dei diritti (in-copyright o in pubblico dominio), i permessi di utilizzazione e le informazioni sui distributori; è utile a biblioteche, archivi e musei che espongono dati sul web e devono indicare i termini di utilizzo delle risorse. CC REL propone cinque classi di entità: proprietà dell'opera, proprietà delle licenze, permessi, requisiti e vincoli (d'uso, di distribuzione, etc.).

g) Ontologie per le *citazioni*

- *Bibliographic Ontology* (BIBO):³² è utilizzato per citazioni di opere e autori; usa alcuni elementi di altri schemi di metadati, tra cui Dublin Core. Definisce, nelle sue specifiche, circa sessanta tipologie di contributo, tra cui atti di convegno, capitoli di libro, film. Interessanti le parti dedicate al trattamento degli autori, allo status del contributo (peer-reviewed, accettato, rifiutato), e all'ISBN, ISSN e altri numeri di controllo.
- *Semantic Publishing and Referencing* (SPAR):³³ è una suite di ontologie differenti e complementari, che possono essere utilizzate singolarmente o in combinazione; è composta di otto ontologie, costruite con linguaggio OWL:
 - FaBIO(FRBR-aligned Bibliographic Ontology)
 - PRO (Publishing Roles Ontology)
 - PSO (Publishing Status Ontology)
 - PWO (Publishing Workflow Ontology)
 - CiTO(Citation Typing Ontology)
 - BiRO(Bibliographic Reference Ontology)

³¹ <<http://creativecommons.org/ns>>.

³² <<http://bibliontology.com>>.

³³ <<http://purl.org/spar>>.

- C4O (Citation Counting and Context Characterization Ontology)
- DoCO (Document Component Ontology)

Queste ontologie sono utili per citazioni e riferimenti bibliografici, per esprimere gli stati di pubblicazione di una risorsa e per strutturare flussi del lavoro editoriale.

h) Ontologie per i *metadati di conservazione*

- *Preservation Metadata: Implementation Strategies* (PREMIS):³⁴ sviluppato da un gruppo di lavoro dell'OCLC (Online Computer Library Center) e di RLG (Research Libraries Group) è strutturato con linguaggio OWL e contiene gli elementi necessari a codificare i dati di conservazione di una risorsa.
- *PRONOM Vocabulary*:³⁵ il vocabolario è stato prodotto dal Digital Preservation Department degli UK National Archives per le descrizioni relative all'attività di conservazione di risorse digitali.

i) Ontologie e vocabolari per i *metadati dei dataset*, oltre al Dublin Core già citato:

- *Vocabulary of Interlinked Datasets* (VoID):³⁶ un vocabolario che consente di rappresentare metadati sui dataset RDF, in particolare, per caratterizzare le informazioni riguardanti le connessioni tra diversi dataset, come, per esempio, i linkset.
- *Data Catalog Vocabulary* (DCAT):³⁷ un vocabolario standardizzato dal W3C utilizzato per metadattare i cataloghi di dati, i dataset contenuti in un catalogo e le diverse modalità di reperimento e accesso ai dati. Attualmente è il vocabolario utilizzato dal catalogo italiano dati.gov.it.
- *Asset Description Metadata Schema* (ADMS):³⁸ un modello utilizzato per rappresentare asset di interoperabilità semantica. Possibili asset sono, per esempio, le classificazioni, i

³⁴ <www.loc.gov/standards/premis/owlOntology-announcement.html>.

³⁵ <www.nationalarchives.gov.uk/PRONOM/Default.aspx>.

³⁶ <<http://www.w3.org/TR/void/>>.

³⁷ <<http://www.w3.org/TR/vocab-dcat/>>.

³⁸ <<http://www.w3.org/ns/adms>>.

modelli, gli standard, etc. che possono essere utilizzati per abilitare l'interoperabilità semantica. Questa ontologia ha l'obiettivo di favorire la scoperta e la federazione di diversi repository di asset semantici.

- PROV:³⁹ un framework standardizzato dal W3C per la rappresentazione della provenienza di artefatti. Consente di modellare il processo di generazione di un artefatto in maniera quasi analoga ai sistemi di controllo versione.
- *Vocabulary for ANNotating vocabulary descriptions* (VANN):⁴⁰ un vocabolario che consente di rappresentare metadati e annotazioni su vocabolari; è utilizzato per arricchirli con esempi e note di utilizzo.
- *Vocabulary of a Friend* (VOAF):⁴¹ un vocabolario per descrivere i vocabolari in linked data e le loro relazioni. Aggiunge, per ciascun vocabolario, i dataset che lo utilizzano.

3.3.2 Cross-mapping tra ontologie

Una tematica aperta è la mappatura di vocabolari e ontologie che attengono allo stesso dominio, ma che usano terminologia differente per esprimere il medesimo concetto.⁴² In ambito catalografico, per esempio, si usano:

- ISBD
- FRBR
- RDA

³⁹ <<http://www.w3.org/TR/prov-overview/>>.

⁴⁰ <<http://vocab.org/vann/.html>>.

⁴¹ <<http://lov.okfn.org/dataset/lov/vocabs/voaf>>.

⁴² L'attività di mappatura trasversale tra ontologie è stata analizzata nello studio per i dati culturali della Provincia autonoma di Trento.

Esempio di dichiarazione di *luogo di pubblicazione* nei silos RDF di ISBD, FRBR, RDA:

- ISBD: luogo di pubblicazione, produzione, distribuzione
- FRBR: luogo di pubblicazione o distribuzione
- RDA: luogo di pubblicazione
- RDA: luogo di distribuzione
- RDA: luogo di manifattura

ISBD combina in un'unica formulazione tre concetti diversi: luogo di pubblicazione, luogo di produzione, luogo di distribuzione;

FRBR combina in un'unica formulazione due concetti diversi: luogo di pubblicazione, luogo di distribuzione;

RDA suddivide i tre concetti di luogo di pubblicazione, di luogo di produzione, di luogo di distribuzione in tre formulazioni corrispondenti a tre elementi diversi.

Il *cross-mapping*, inteso come una sovra-mappatura dei differenti vocabolari, è una soluzione per riportare all'unicità formulazioni diverse di medesimi concetti o entità, tramite la strutturazione di relazioni di equivalenza tra le varie dichiarazioni:

FRBR:LuogoDiPubblicazioneODistribuzione *sameAs*

RDA:luogoDiPubblicazione *sameAs* RDA:luogoDiDistribuzione

L'agente che intercetta la formulazione FRBR comprende che i concetti "luogo di pubblicazione o distribuzione" equivalgono a due elementi differenti espressi da RDA (luogo di pubblicazione, luogo di distribuzione).

4. SEMANTIC WEB STACK

I linked data sono definibili, dunque, come *una tecnologia* del web semantico. La realizzazione del web semantico comporta, tuttavia, l'uso di un insieme di standard e strumenti che può essere rappresentato come una *pila tecnologica*, cioè come una sequenza di livelli sovrapposti ciascuno dei quali comprende componenti tecnologici che assolvono a specifiche funzioni, comunemente identificata con le formulazioni *Semantic web stack* (letteralmente: pila del web semantico), *Semantic web cake* o *Semantic web layer cake*. Essa rappresenta l'architettura del web semantico.¹

Il Semantic web stack è un esempio della gerarchia degli standard e delle tecnologie utilizzate per la costruzione del web semantico, in cui ogni strato sfrutta e usa le capacità degli strati sottostanti. La pila tecnologica, infatti, va letta dal basso verso l'alto; ciascuno strato ha bisogno del precedente. Lo schema mostra che il web semantico è un'estensione (non una sostituzione) del web ipertestuale.² L'immagine riprodotta alla pagina seguente è stata creata da Tim Berners-Lee per visualizzare schematicamente la disposizione a strati dei componenti che la concretizzano e ne consentono la realizzazione. Esistono diverse rappresentazioni della pila che esprimono la maggiore o minore complessità tecnologica.

Il Semantic web stack è utile al lettore perché la raffigurazione offre una visione complessiva delle possibili tecnologie utilizzate nella realizzazione semantica.

Le due immagini di stack che seguono presentano il medesimo livello di dettaglio, divergono soltanto per la rappresentazione grafica.

¹ <<http://stack.lod2.eu/blog/>>.

² Vedi: Alberto Salarelli, *Sul perché, anche nel mondo dei linked data, non possiamo rinunciare al concetto di documento*, "AIB studi", vol. 54, n. 2/3 (maggio/dicembre 2014), p. 279-293.

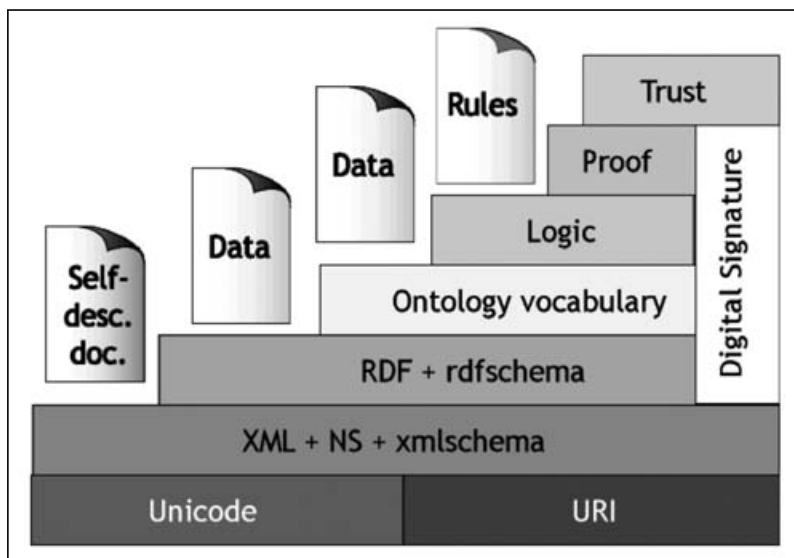


Figura 22 – Immagine originaria del Semantic web stack, creata da Tim Berners-Lee nel 2008

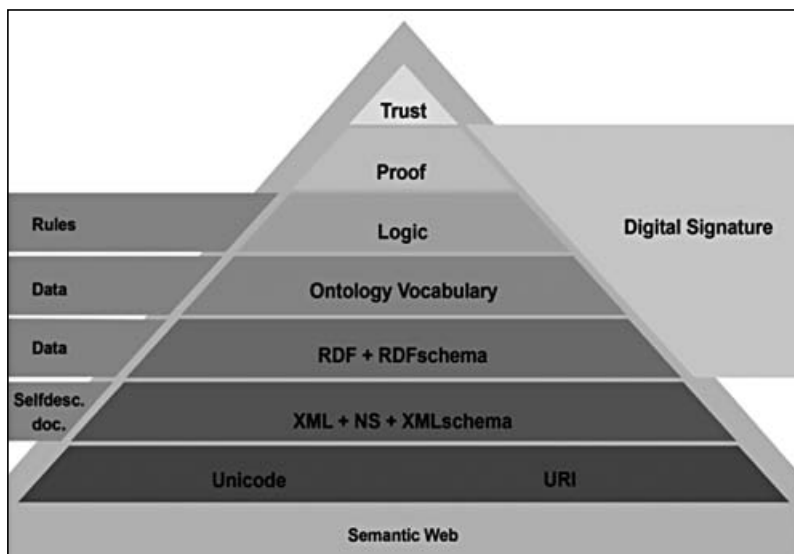


Figura 23 – Immagine variante del Semantic web stack (<www.dbis.informatik.uni-frankfurt.de/~tolle/RDF/index_e.html>)

La terza immagine di stack opera una suddivisione diversa rispetto alle due precedenti e introduce ulteriori tecnologie e standard per l'implementazione (attuazione pratica), per esempio, il linguaggio SPARQL e lo strato superiore che comprende le interfacce e le applicazioni che consentono la fruizione del web semantico da parte degli utenti finali e di altre applicazioni e servizi di rete.

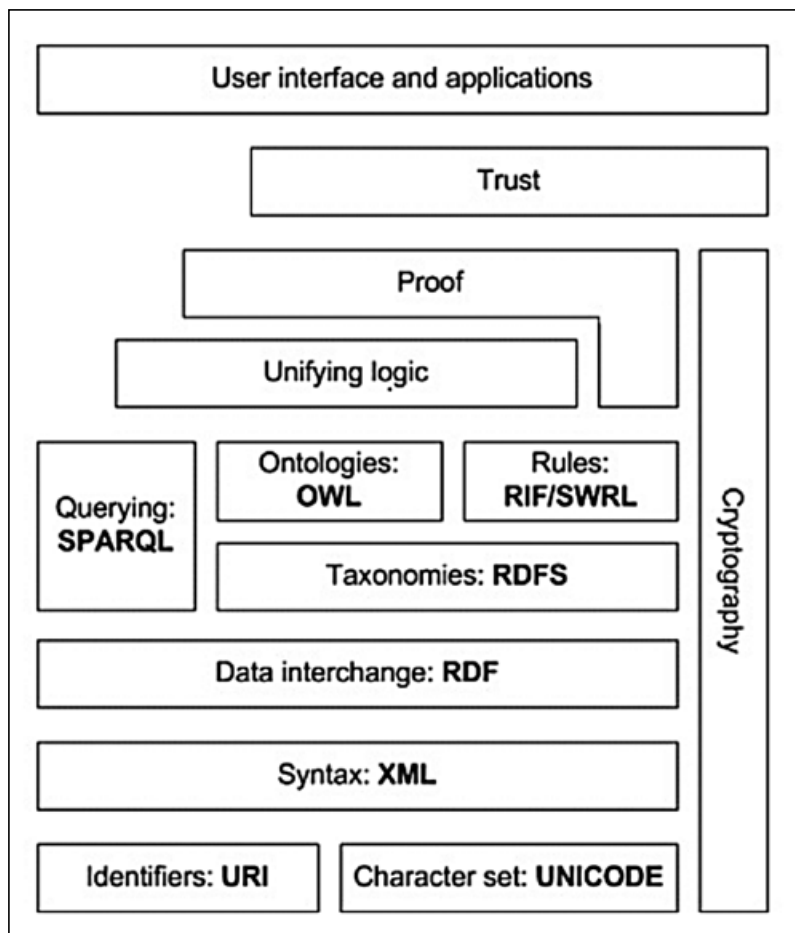


Figura 24 – Immagine variante del Semantic web stack

La quarta rappresentazione di stack riprodotta, disegnata da Tim Davis nel 2011, amplia la descrizione del livello di fruizione (using), ponendo in evidenza:

- a) l'importanza dei motori di ricerca funzionalmente estesi per derivare informazioni dai linked data;
- b) l'utilizzo della tecnologia mashup per combinare i dati provenienti da fonti diverse e tramite questi creare nuove applicazioni;
- c) come i linked data possono contribuire a migliorare i livelli di produttività in ambiti quali la ricerca e la business intelligence, grazie alla maggiore integrazione dei dati a livello semantico.

Questo stack comprende un livello trasversale relativo alla fruibilità dei linked data in modo aperto (open linked data) che a sua volta include un aspetto di tipo normativo, la definizione di licenze per regolare l'uso di questi dati e una componente tecnologica relativa all'adozione di formati di scambio aderenti a standard aperti.

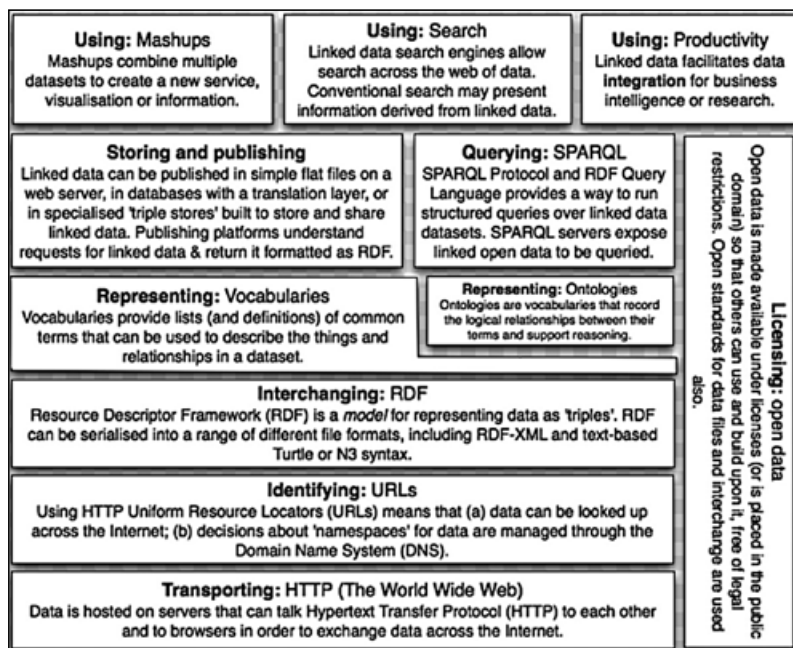


Figura 25 – Immagine variante del Semantic web stack disegnata da Tim Davies nel 2011

La quinta e ultima immagine di stack, disegnata da Benjamin Nowack nel 2011, evoluzione cronologica delle precedenti, presenta livelli maggiormente dettagliati, tra cui XML Turtle e RDFa, che sono sintassi diverse per poter scrivere un grafo RDF in formato testuale. La rappresentazione tridimensionale mette in evidenza la dimensione funzionale. Nel piano, l'asse delle ascisse riporta i diversi standard di riferimento, mentre quello delle ordinate indica i corrispondenti concetti astratti del modello; l'asse verticale descrive la stratificazione dei vari livelli funzionali, sempre secondo il modello dal basso verso l'alto (bottom-up), e colloca gli standard e gli strumenti nell'appropriato ambito funzionale. La rappresentazione pone bene in evidenza due aspetti: il primo riguarda le tecnologie del web (web platform), che sono il fondamento su cui si basa il web semantico; il secondo è relativo al cilindro che mostra come i linked data usano solo una parte degli standard e delle tecnologie del più complesso stack necessario per le realizzazioni del web semantico.

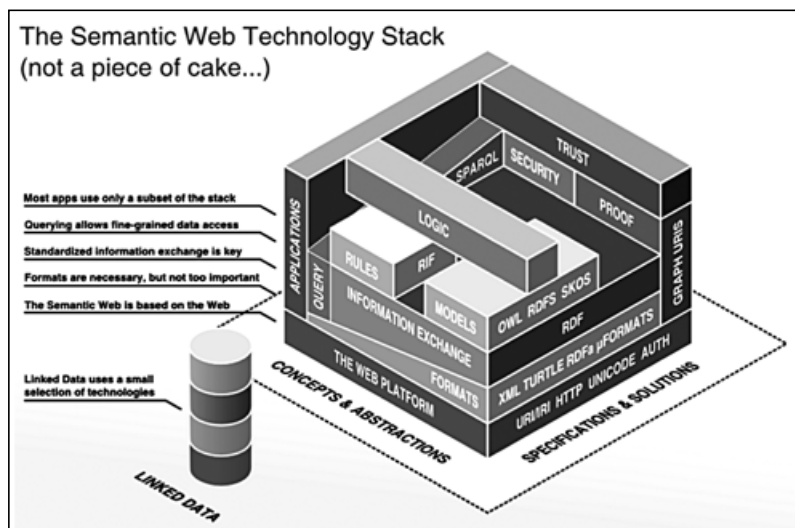


Figura 26 – Immagine tridimensionale del Semantic web stack – o cake – disegnata da Benjamin Nowack nel 2011

Tra gli esempi di Semantic web stack possiamo citare la piattaforma open source sviluppata in ambito Apache.org, denominata Marmotta <http://marmotta.apache.org/>, che comprende:

Read write linked data: servizi di lettura di dati o d'interrogazione (query) e di scrittura per l'aggiornamento, l'integrazione e la cancellazione (update, delete) del dataset.

RDF store with transaction, versioning and rule-based reasoning: RDF store che permette di memorizzare triple; supporta:

- transazioni in termini di atomicità, integrità, durabilità e consistenza dei dati;
- versionamento del dataset;
- processi inferenziali basati su regole.

Transparent linked data caching: possibilità di utilizzare, in maniera trasparente, servizi di cache (memorizzazione temporanea) per ottimizzare i tempi di risposta delle query.

Integrated basic security mechanism: include meccanismi di sicurezza di tipo basilare, per proteggere l'accesso ai servizi di query e per proteggere l'aggiornamento del dataset.

La figura che segue descrive l'architettura tecnico-funzionale e i componenti di Marmotta.

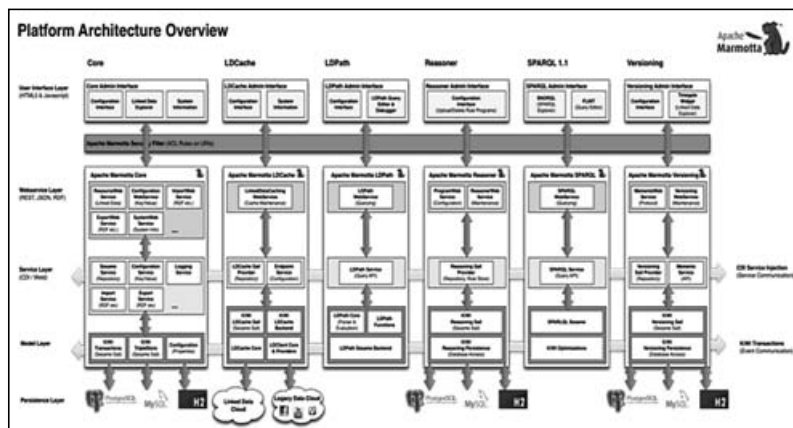


Figura 27 - Architettura della piattaforma Marmotta, <<http://marmotta.apache.org/platform/>>

RDF store (kiwi) implementa la persistenza delle triple RDF avvalendosi dell'utilizzo di un DB relazionale (H2, PostgreSQL o MySQL). Lo strato dei servizi interagisce con quello volto all'utilizzo tramite una serie di web services specializzati.

5. PUBBLICARE LINKED DATA NEL WEB

Lo scenario di spazio globale rappresentato dal web, in cui le informazioni sono lette, manipolate e fruite da agenti umani ed elettronici, genera un nuovo atteggiamento in chi produce fonti informative: se in passato l'accento era sulle esigenze locali o riferite a una comunità specialistica, adesso ogni decisione dovrebbe essere assunta considerando il contesto globale di condivisione. Ciò che viene prodotto potrà essere usato da chiunque in qualsiasi comunità e in qualsiasi contesto; l'orizzonte è il mondo intero. L'interoperabilità tra sistemi è un valore acquisito ed è tanto più possibile ed economico quanto più i dati prodotti sono strutturati e ricchi. Scelte locali e particolari risultano antieconomiche, non potendo sfruttare la ricchezza che il web propone con i contributi di altre comunità, e condannano a un isolamento penalizzante, sia per lo spreco di risorse e di tempo, sia per il mancato riuso del dato prodotto e, dunque, per la sua inutilità. Occorre evitare l'emarginazione delle biblioteche e degli istituti culturali in uno spazio apparentemente aperto e pubblico (*essere nel web*), ma non realmente integrato e condivisibile (*essere del web*).

Esistono due modi per costruire il web semantico: aggiungendo struttura alle pagine HTML esistenti; nelle biblioteche ciò equivale ad agire sulle interfacce dei cataloghi ormai diffusamente fruibili via web, o producendo dati strutturati in linked data; il secondo approccio ha un impatto più profondo sulle tradizionali modalità di catalogazione e, soprattutto, sui software gestionali delle biblioteche che sono basati sugli standard di tipo MARC, ma trova una perfetta corrispondenza nell'adozione del nuovo standard di catalogazione RDA. Nell'uno e nell'altro caso si tratta di aggiungere significato alle informazioni, in una modalità interpretabile dalle macchine oltre che da persone e gruppi. Pubblicare dati nel web è relativamente facile e i vincoli tecnologici sono davvero pochi e semplici; tuttavia, affinché l'informazione sia *del web* e non solo *nel web* è necessario rispettare alcuni principi e alcune prescrizioni tecnologiche: più si rispettano questi principi e queste prescrizioni più il dato pubblicato è

considerato *di qualità*, secondo un modello rappresentato dalle 5 stelle di Tim Berners-Lee.¹

5.1 Le 5 stelle di Tim Berner-Lee

Questo sistema di valutazione dei dati pubblicati nel web esprime un metro di giudizio imperniato su 5 requisiti, che rappresentano 5 differenti livelli incrementali basati su caratteristiche di *apertura* e di *collegamento* dei dati nel web.

Ogni stella conquistata equivale a un dato più condivisibile, comprensibile e riusabile nel web.

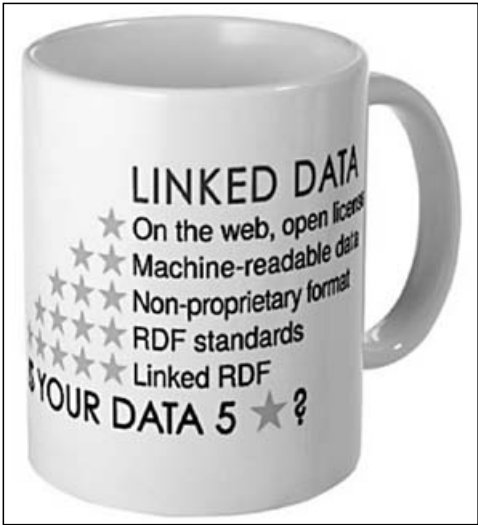


Figura 28 – Le 5 stelle dei linked data

★	Dati disponibili nel web (in qualunque formato), ma con una licenza aperta, per essere <i>open data</i> .
★★	Dati disponibili come dati strutturati in formati leggibili da una macchina (per esempio, excel, anziché un'immagine che riproduce una tabella).
★★★	Come il livello precedente, ma con un formato non proprietario (per esempio, CSV, anziché excel).
★★★★	Tutto quello di sopra, ma usando gli standard del W3C (RDF e SPARQL) per identificare le cose, così che le persone possano puntare al tuo prodotto.
★★★★★	Tutto quello di sopra, più: connettere i tuoi dati a quelli di altre persone per fornire il contesto.

¹ Vedi la sezione *Is your linked open data 5 star?*, <www.w3.org/Design/Issues/LinkedData.html>.

5.2 La qualità dei dati pubblicati

La valutazione dei dati pubblicati, secondo il modello di Tim Berners-Lee, considera cinque punti fondamentali (ciascun punto successivo si aggiunge al precedente).

1. La disponibilità dei dati nel web in qualsiasi formato, purché pubblicati con licenze *open*. La tematica delle licenze d'uso e di distribuzione è un tema aperto di riflessione tra produttori e distributori di risorse elettroniche. Le formule finora trovate per la distribuzione delle risorse nel web sono spesso complesse e combinano differenti parametri (distribuzione, riproduzione, uso, vendita, etc.). Una tipologia di licenza molto utilizzata nel web sono le licenze *Creative Commons* (CC).²
2. La strutturazione dei dati in modalità leggibile da una macchina, preferendo per esempio, una tabella in Excel a una scansione dell'immagine della tabella: ciò garantisce una migliore e più diffusa usabilità del dato da parte delle macchine, anche se comporta una minore leggibilità da parte dell'uomo.
3. La scelta di formati non proprietari (per esempio, in CSV, Comma Separated Values, anziché in Excel, che è un formato proprietario). I vantaggi nell'impiego di formati open sono numerosi: costo contenuto della realizzazione del progetto, livello alto d'interoperabilità tra applicativi, sicurezza elevata rispetto ad attacchi di virus, garanzie solide per la conservazione a lungo termine delle risorse digitali; e, inoltre, la presenza di una *comunità di sviluppatori* del software open source che dovrebbe garantire una minore dipendenza dalla società *proprietaria* del software commerciale e maggiori possibilità che il software continui a essere supportato ed evoluto nel tempo.
4. L'utilizzo di standard raccomandati dal W3C (come RDF o SPARQL) per identificare gli oggetti, in modo che gli agenti elettronici e umani possano puntare inequivocabilmente e

² Vedi: <<http://creativecommons.org/>>. Europeana utilizza la licenza CC; vedi: <<http://pro.europeana.eu/linked-open-data>>.

stabilmente a queste risorse. La migliore strutturazione per gli open linked data è garantita dai formati di serializzazione RDF e, dunque, RDF in una sintassi prevista: RDF/XML, RDFa, N-Triples, Terse RDF Triple Language Turtle.

5. Il collegamento interno ed esterno dei dati (link interni e interlinking) con dati prodotti da altri attori, in modo da definire un contesto che superi l'ambito del dataset locale. Il meccanismo dell'interlinking consente un ampliamento dello spazio da una dimensione locale e circoscritta a una dimensione globale.

A proposito del rapporto tra *open data*, *linked data* e *web semantico* è interessante l'osservazione fornita da Paul Walk.³

- 1) i dati possono essere *open*, ma **non** essere *linked*;
- 2) i dati possono essere *linked*, ma **non** essere *open*;
- 3) i dati *open e linked* sono in continuo arricchimento;
- 4) il *web semantico* può funzionare solo con dati *open e linked*.



Figura 29 – Le 5 stelle dei linked open data

Più *open data* saranno *connessi* nel web con un significato ben definito, più le applicazioni web saranno performanti.

³ Paul Walk's weblog, 2009: <<http://blog.paulwalk.net/2009/11/11/linked-open-semantic/>>.

5.3 I livelli di qualità in relazione a informazione, accesso e servizi

Le *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico* del 2014, emanate dall'Agenzia per l'Italia digitale della Presidenza del Consiglio dei ministri,⁴ al paragrafo 4.1, *Modello per i dati*, riprendono la tematica per ricondurla a tre livelli d'impiego: *informazione*, *accesso* e *servizi* implementabili: si tratta di un arricchimento del modello a 5 stelle, che fornisce ulteriori elementi per comprendere il livello di qualità offerto ai fini del loro massimo riutilizzo. Le sigle del riquadro a sinistra esprimono le caratteristiche dei dati che, variamente composte, corrispondono ai cinque gradi di qualità:

OL: licenze aperte

RE: formati leggibili dalle macchine

OF: formati aperti (non proprietari)

URI: utilizzo di identificativi URI

LD: dati collegati

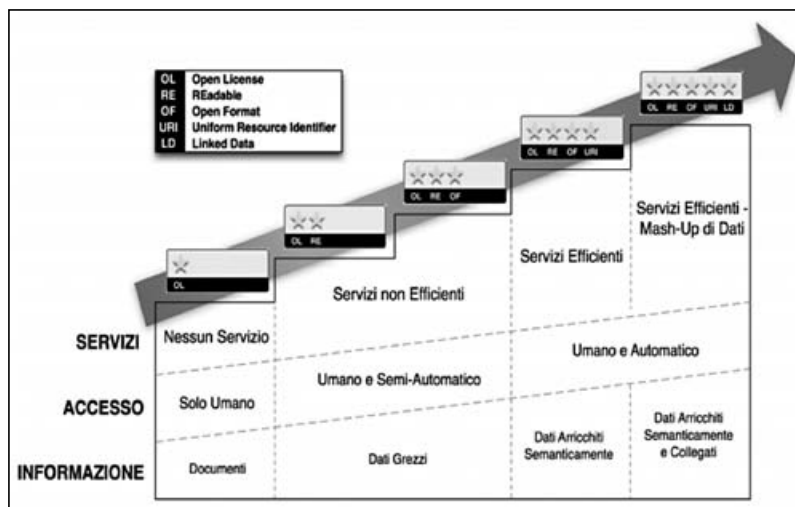


Figura 30 – Schema dei livelli, accesso e informazione dei dati in relazione alle 5 stelle

⁴ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_publicolg2014_v0.7finale.pdf>.

Più nel dettaglio, i cinque livelli sono:



OL: licenze aperte

Dati disponibili in qualunque formato, ma con una licenza aperta. I dati possono rappresentare documenti, essere accessibili solo dagli agenti umani e non offrire alcun servizio.



OL: licenze aperte
RE: formati leggibili dalle macchine

Dati disponibili in un formato leggibile da un agente automatico. Tipicamente, rientrano in questo livello dati in formati proprietari (per esempio, Excel), che possono essere, però, letti (REadable) e interpretati anche da un agente automatico).



OL: licenze aperte
RE: formati leggibili dalle macchine
OF: formati aperti (non proprietari)

Dati con caratteristiche del livello precedente ma con un formato non proprietario (per esempio, CSV, JSON). I dati possono avere un'interpretazione sia umana sia automatica. Ciò nonostante, questi dati non hanno ancora le caratteristiche sufficienti per fornire una base solida in grado di fornire servizi efficienti.



OL: licenze aperte
RE: formati leggibili dalle macchine
OF: formati aperti (non proprietari)
URI: utilizzo di identificativi URI

Dati con caratteristiche del livello precedente, ma esposti usando gli standard W3C RDF e SPARQL. Questi dati possono esse-

re arricchiti semanticamente, letti sia dagli agenti umani sia dagli agenti automatici, e costituiscono una ricca base per poter creare servizi efficienti.



OL: licenze aperte

RE: formati leggibili dalle macchine

OF: formati aperti (non proprietari)

URI: utilizzo di identificativi URI

LD: dati collegati

Dati con caratteristiche del livello precedente ma connessi a dati esposti da altre persone e organizzazioni. I dati di questo livello vengono potenziati dai meccanismi di *mashup* , un particolare tipo di applicazione web, che consente l'integrazione, a differenti livelli, di servizi, o API, messi a disposizione in rete. I mashup sono costituiti da componenti che, sincronizzati e debitamente orchestrati (modellati), consentono l'integrazione di contenuti diversi, generando servizi efficienti. I dati di questo livello possono essere considerati sufficientemente modellati per dare luogo alla pubblicazione di LOD: dati aperti e collegati ad altri dati esposti nel web da una varietà di agenti. È un metodo efficace per risolvere problemi d'identità, provenienza, semantica, integrazione e interoperabilità.

5.4 Ciclo di vita dei linked data

La pubblicazione di linked data prevede un workflow di lavorazione che parte con la selezione di ciò che si vuole rendere disponibile.⁵ Il processo è stato variamente illustrato, ma le fasi proposte nei differenti schemi differiscono poco l'una dall'altra; il flusso logico e pratico per arrivare al risultato è, infatti, ormai chiaro, consolidato e condiviso dalle comunità che si occupano di linked data e linked open data. Il W3C propone

⁵ Vedi: *First Workshop on Managing the Evolution and Preservation of the Data Web*, <<http://www.diachron-fp7.eu/workshops.html>>, co-located with ESWC 2015, Portorož, Slovenia.

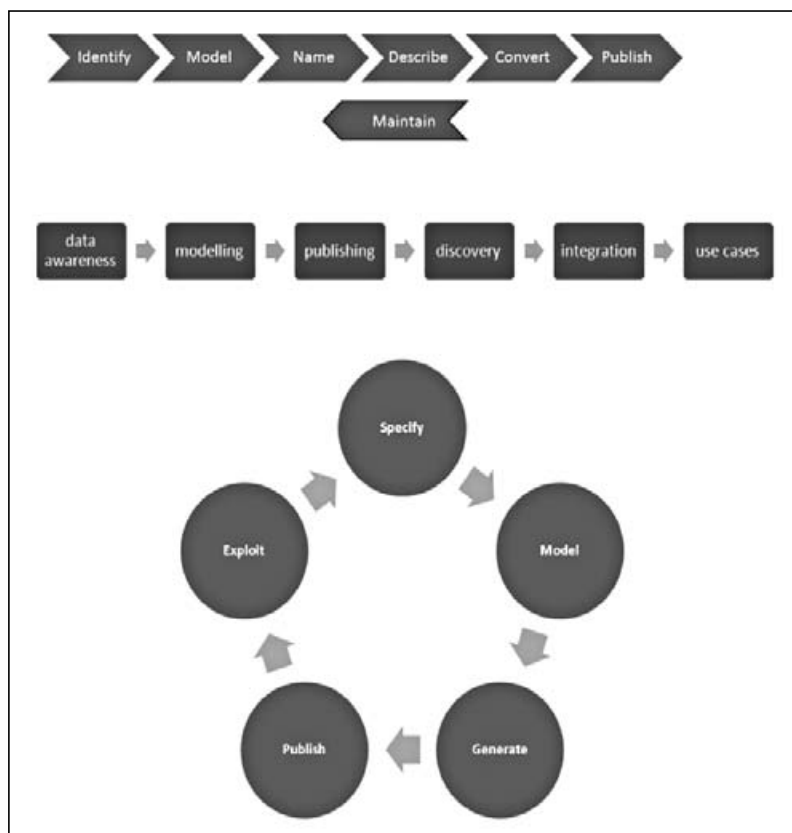


Figura 31 – <<http://www.w3.org/TR/ld-bp/>>

un prontuario con le buone pratiche per pubblicare i linked data⁶ rivolte principalmente ai dati governativi (tra i primi a essere stati pubblicati come open), estendibile ai dati prodotti da ogni altro attore. Lo schema di Boris Villazón-Terrazas, Luis. M. Vilches-Blázquez, Oscar Corcho e Asunción Gómez-Pérez, *Methodological guidelines for publishing linked data* del 2011,⁷ è divenuto un riferimento, citato ampiamente nella letteratura,

⁶ <<https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>>.

⁷ <https://www.lri.fr/~hamdi/datalift/tuto_inspire_2012/Suggested_readings/egovld.pdf>.

che si sviluppa in 5 passi (talora 7 o anche 8 passi), schematizzati in grafi differenti, sequenziali e circolari:

1. Definizione delle specifiche, la fase include:
 - identificazione e analisi della fonte dei dati;
 - disegno dell'URI;
 - definizione delle licenze di distribuzione, uso, etc.
2. Modellizzazione; la scelta o la creazione di ontologie e vocabolari controllati.
3. Generazione dei dati in formato RDF, tramite i linguaggi di mappatura disponibili; questa fase di generazione include:
 - conversione del dato;
 - bonifica di quanto prodotto;
 - creazione di collegamenti, come la creazione di link manuali o in modalità automatica.
4. Pubblicazione dei dati in RDF:
 - pubblicazione del dataset su uno o più portali o piattaforme di pubblicazione (per esempio, CKAN Registry, Comprehensive Knowledge Archive Network⁸ e Socrata⁹) e dei relativi metadati;¹⁰
 - generazione di sistema per aumentare la ricercabilità del dato.
5. Valorizzazione dei dati tramite la costruzione di strumenti semplici ed efficaci di utilizzo del dataset; per esempio, tramite la costruzione di interfacce per la fruizione.

Esistono altri modelli per il workflow di pubblicazione dei linked data; Bernadette Hyland, per esempio, propone un flusso di produzione articolato in sette passi:

1. identificazione della fonte dei dati;
2. modellizzazione del vocabolario;
3. definizione nelle logiche di creazione degli URI;
4. descrizione delle risorse, tramite l'utilizzo di linguaggi e strumenti (tool) aperti;

⁸ CKAN – The Data Hub, *Linking open data cloud*, <<http://thedatahub.org/group/lodcloud>>, 2012; sito italiano, <<http://it.ckan.net>>.

⁹ <<http://www.socrata.com/>>.

¹⁰ Sito italiano, <<http://it.ckan.net>>.

5. conversione nei formati previsti dalla tecnologia dei linked data;
6. pubblicazione dei dati nella modalità più aperta e trasparente possibile;
7. manutenzione del dataset per scongiurare l'obsolescenza o l'indisponibilità del dato.

I modelli servono a evidenziare una logica unitaria e condivisa, guidata dalla chiarezza dell'obiettivo dell'elaborazione dei dati, nonché dalla semplicità delle tecnologie e dei linguaggi utilizzati.¹¹

5.5 Metadatazione del dataset

La metadatazione del dataset ricopre un ruolo essenziale in fase di pubblicazione, essendo questa la fase che espone i dati a terze parti: essa fornisce informazioni importanti sul dataset pubblicato, tra cui:

- *informazioni generali*: il nome del dataset, il creatore, l'editore, la data di pubblicazione e di aggiornamento, etc.
- *dati d'accesso e utilizzo* del dataset: informazioni sugli http URI, sull'esistenza di SPARQL endpoint, sulla disponibilità del dataset come dump, etc.
- *dati strutturali*: il namespace del dataset, i vocabolari utilizzati per la conversione, dati statistici riferiti al dataset etc.
- *dati sui link esistenti tra vari dataset*: i linkset, il tipo di link che collega due differenti dataset, etc.

I metadati, dunque, consentono una maggiore comprensione del dataset RDF e aiutano, pertanto, la ricerca e il riutilizzo dei dati stessi. Esistono diversi modelli utili per la metadatazione dei dataset,¹² alcuni dei quali si focalizzano principalmente sugli aspetti di schema (per esempio, Dublin Core) o sul formato.

¹¹ Per approfondimenti vedi il progetto LOD2, *The life-cycle comprises in particular the stages*, <<http://stack.lod2.eu/blog/>>.

¹² Si veda nel paragrafo 3.3.1 la sezione dedicata alle ontologie e vocabolari per i metadati.

Nelle citate *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico* del 2014, dell'Agenzia per l'Italia digitale della Presidenza del Consiglio dei ministri,¹³ al paragrafo 4.2, *Modello per i metadati* viene proposto uno schema di classificazione della qualità dei metadati rivolto soprattutto allo sviluppo di servizi, quindi all'accessibilità dei metadati e al loro livello di granularità. La classificazione qualitativa del modello è basata su due fattori:

- il *legame tra dato-metadato*: questa dimensione indica quanto i metadati riescono a essere, fisicamente e logicamente, legati ai dati anche dopo un possibile processo di trasformazione e utilizzo:
 - *legame assente*: i dati non sono accompagnati da un'opportuna metadatazione (livello 1);
 - *legame debole*: i dati sono accompagnati da metadati esterni, per esempio, da schede nella pagina di download o in file separati (livello 2);
 - *legame forte*: i dati incorporano i metadati che li descrivono (livello 3 e 4).
- il *livello di dettaglio dei metadati*: questa dimensione rappresenta l'ambito dei metadati, ovvero la granularità della loro descrizione:
 - *nessuno*: i dati non sono accompagnati da un'opportuna metadatazione (livello 1);
 - *dataset*: i metadati forniscono informazioni relativamente a un dataset, sono informazioni condivise dall'insieme di dati interni a quel dataset (livelli 2 e 3);
 - *dato*: i metadati forniscono informazioni relativamente al singolo dato, quindi col massimo grado di dettaglio possibile (livello 4).

¹³ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicolg2014_v0.7finale.pdf>.

Di seguito, lo schema a quattro livelli, rappresentativo delle combinazioni ammissibili dei gradi di differenti fattori:

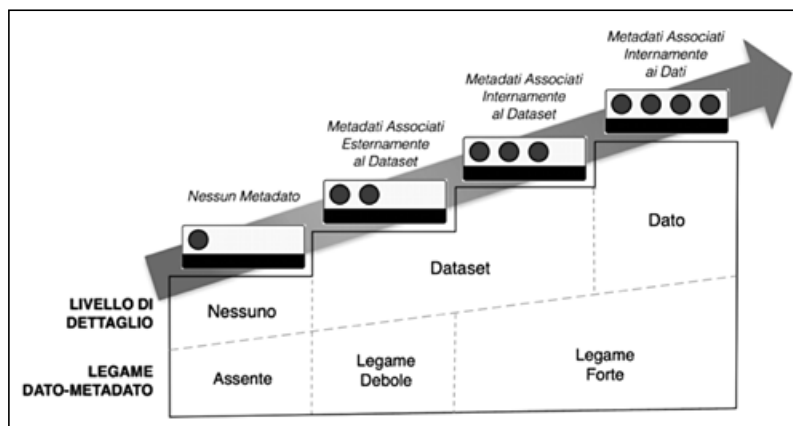


Figura 32 – Modello per la definizione dei metadati

La possibilità di raggiungere un livello di metadatatizzazione dipende da diversi fattori, tra cui la natura dei dati e il loro grado di granularità. Il livello 4, ove possibile, è certamente l'ottimale.

5.6 Linking Open Data Project

Il Linking Open Data Project,¹⁴ nato per volere della comunità dei linked open data, ha lo scopo di favorire la pubblicazione nel web di dataset RDF e di semplificare le operazioni di interlinking tra risorse differenti. Il progetto condivide un forum per la partecipazione alle attività che la pubblicazione di grandi quantità di dati necessariamente richiede, tra cui:

- browsing e altre funzioni di ricerca delle risorse (per compiere indagini sui dati disponibili);
- semantic web crawler e motori di ricerca (per interrogare i dataset);
- repository RDF e motori inferenziali;
- strumenti di pubblicazione dei linked data;

¹⁴ <<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>>.

- algoritmi di creazione dei collegamenti tra risorse;
- strumenti per le bonifiche dei dati.

Il progetto intende velocizzare le attività di pubblicazione dei dataset nel web, coinvolgendo il maggior numero possibile di persone e cercando di risolvere i problemi semplici rimandando quelli complessi. Il motto è “pubblicare adesso e rifinire poi”. Esempio di questo impegno condiviso è la pubblicazione del dataset Dbpedia che ha comportato:

- l'estrazione di informazioni strutturate da *Wikipedia*;
- l'apertura dei dati nel web tramite licenze open;
- la creazione di link tra il dataset DBpedia e altri dataset disponibili nel web.

Per meglio esprimere la dimensione delle informazioni ottenute da questa operazione è utile riferire alcuni dati:

- 1.600.000 concetti (di cui 58.000 persone, 70.000 luoghi, 35.000 album musicali, 12.000 film);
- 91 milioni di triple con 8141 proprietà diverse;
- abstract in 10 lingue diverse;
- 557.000 collegamenti alle immagini;
- 1.300.000 link a pagine web esterne rilevanti;
- 207.000 categorie di *Wikipedia*;
- 75.000 categorie YAGO⁴⁴

I maggiori collaboratori del progetto sono:

- Freie Universität di Berlino
- Universität Leipzig
- Open Link Software (UK)
- Linking Open Data Community (W3C SWEO).

La ricchezza di questo dataset consiste nel numero di concetti trattati e, soprattutto, nel grandissimo numero di link a dataset esterni; per esempio:

```
<http://dbpedia.org/resource/Berlin>
owl:sameAs
<http://sws.geonames.org/2950159>
```

¹⁵ <[http://en.wikipedia.org/wiki/YAGO_\(database\)](http://en.wikipedia.org/wiki/YAGO_(database))>. I dati si riferiscono alla fine di gennaio 2015.

oppure

<http://dbpedia.org/resource/Tim_Berners-Lee>

owl:sameAs

<<http://www4.wiwiss.fu-berlin.de/dblp/resource/person/100007>>

e nel numero di link da dataset a DBpedia; per esempio:

<<http://richard.cyganiak.de/foaf.rdf#cygri>>foaf:topic_interest

<http://dbpedia.org/resource/Semantic_Web>

oppure

<<http://www.w3.org/People/Berners-Lee/card#i>>

owl:sameAs

<http://dbpedia.org/resource/Tim_Berners-Lee>

L'enorme mole di collegamenti verso l'esterno e dall'esterno ha reso il dataset DBpedia il fulcro da cui parte la rappresentazione dei linked open data pubblicati nel web, immagine conosciuta come *Diagramma dei Linked Open Data cloud* (LOD cloud). Tim Berners-Lee afferma che RDF consente a ciascuno di dire qualsiasi cosa su qualsiasi argomento: AAA: *Anyone can say Anything about Any topic*;¹⁶ l'assunto conduce alla creazione di un grafo di dati globale dalla natura multitematica, multiprospettica e multimediale.

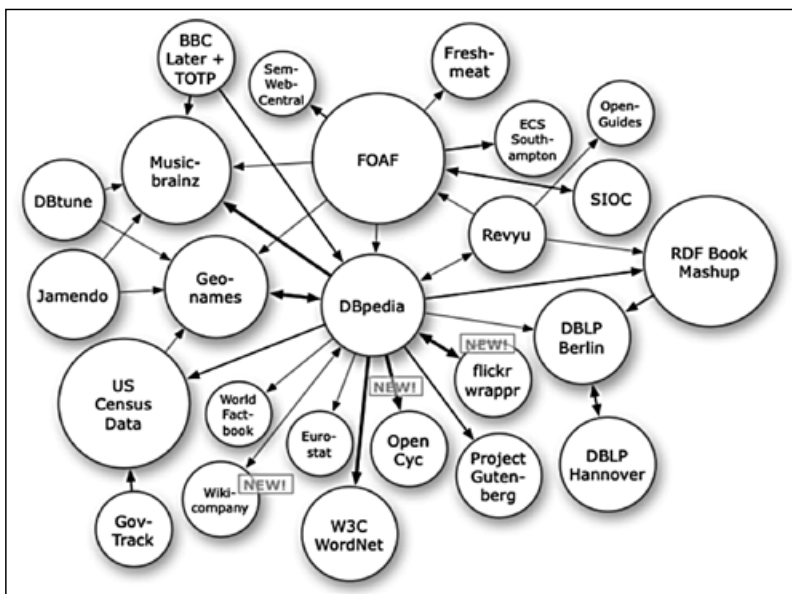
5.7 Linked Open Data cloud

Il progetto Linking Open Data ha registrato un incremento elevatissimo e in pochi anni ha aggregato un gran numero di data-

¹⁶ Il fenomeno di crescita esponenziale di dati nel web ha generato il concetto di Big data: "una raccolta di dataset così grande e complessa da richiedere strumenti differenti da quelli tradizionali, in tutte le fasi del processo: dall'acquisizione, alla curation, passando per condivisione, analisi e visualizzazione" (*Wikipedia*). Il concetto si è diffuso rapidamente ed è stato condiviso da chiunque si occupi di raccolta, analisi, gestione, proprietà e diffusione dei dati, come ricercatori, professionisti, agenzie di produzione, etc. Si veda: Pascal Hitzler, Krzysztof Janowicz, *Linked data, big data, and the 4th paradigm*, <<http://semantic-web-journal.net/system/files/swj488.pdf>>; e, inoltre: No hace falta ser matemático para ser experto en 'big data', <http://economia.elpais.com/economia/2015/02/12/actualidad/1423740778_503816.html>.

set. Di seguito le immagini che illustrano la crescita esponenziale del diagramma, che rappresenta lo *historic landscape*¹⁷ di dataset, compresi i dati chiusi e quelli aperti tramite appropriate licenze e formati,¹⁸ pubblicati nel web come linked (open) data:¹⁹

nel 2007

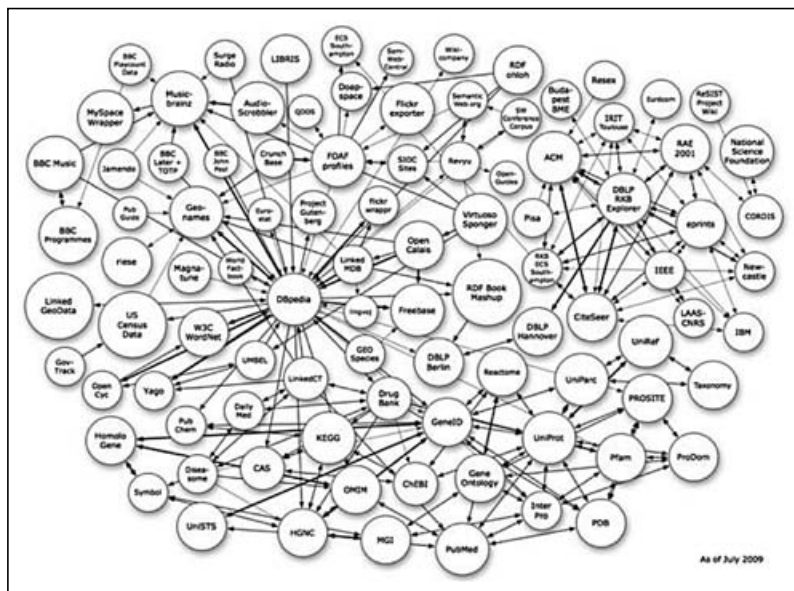


¹⁷ D2.1 Best practice report on cultural heritage linked data and metadata standards, ICT-PSP Project no. 270905 Linked Heritage, November 2011, p. 25, <<http://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&ved=0CEUQFjAB&url=http%3A%2F%2Fwww.linkedheritage.eu%2FgetFile.php%3Fid%3D229&ei=tR3kUvKJOuTeygOZuYGYDw&usg=AFQjCNH2OsFycJ1Mip13aAKt5INSaNAHMA&sig2=K-ATqWM9f5iuMITfnQLt7Q>>.

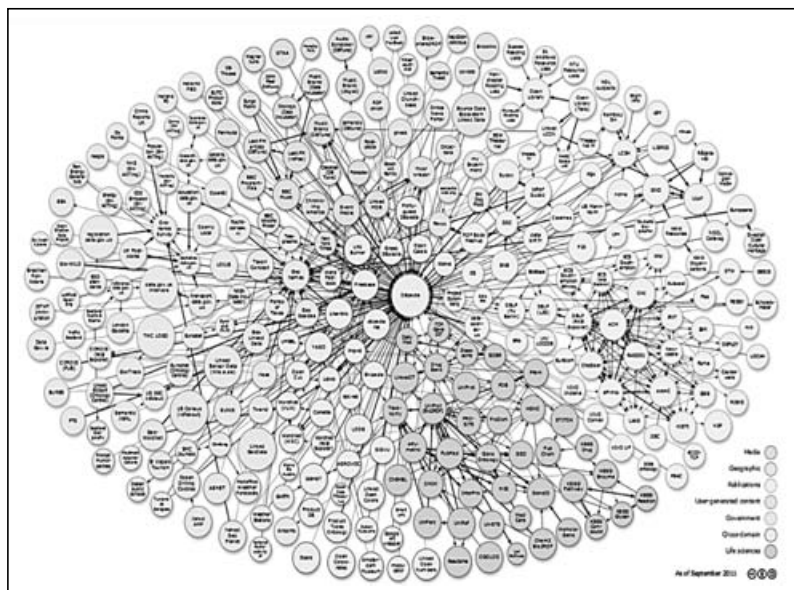
¹⁸ Open Data Charter and Technical Annex del G8 - Policy paper; il principio 1 è: Open Data by default, 2013, <<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>>; Antonella De Robbio, *Forme e gradi di apertura dei dati. I nuovi alfabeti dell'open biblio tra scienza e società*, "Biblioteche oggi", vol. 30, n. 6 (luglio-agosto 2012).

¹⁹ Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch, <<http://lod-cloud.net/>>.

nel 2009



nel 2011



nel 2014

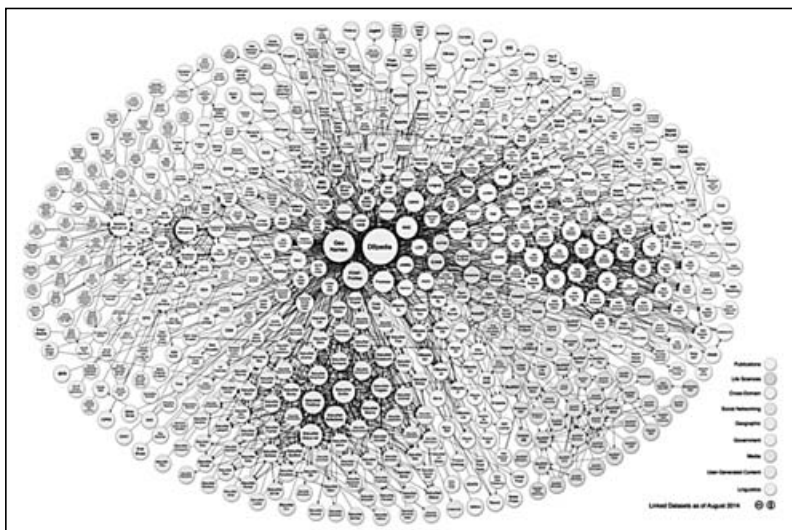


Figura 33 – La nuova versione contiene 570 dataset collegati che sono connessi a 2909 linkset

I *dataset* rappresentati nella *LOD cloud* possono provenire da domini:

1. *privati* (pacchetti di dati chiusi);²⁰
2. *pubblici* (pacchetti di dati aperti tramite formati e licenze aperte; per esempio, *Public Domain Dedication Licence* (PDDL), oppure *Creative Commons Zero Waiver*, o CC0.

Dall'analisi dello stato dell'arte dei linked data compiuta entro il Progetto Linked Heritage risulta che la maggior parte dei pacchetti di dati nella *LOD cloud* non sono aperti (42,6% di dataset aperti vs 57,4% di dataset non aperti).²¹

Nell'ottica di dataset (unità informative e loro parti) esposti *open* tramite la tecnologia dei linked data nel web semantico tre sono i fattori di maggiore interesse:

²⁰ AI3: Adaptive Information, Adaptive Innovation, Adaptive Infrastructure, *What is linked data?*, Blog AI3, posted June 23, 2008, <<http://www.mkbergman.com/447/what-is-linked-data/>>.

²¹ Deliverable D2.1 *Best practice report on cultural heritage linked data and metadata standards*, Linked Heritage Project, cit., p. 25.

1. la fruibilità e il riuso di unità informative (per esempio, dataset di *Open Library*, *British National Bibliography*, *Hungarian National Library*, *CrossRef*, *Europeana*; *Metadata Element Sets*;²² *Linked Open Vocabularies LOV*²³) in una molteplicità di soluzioni di collegamenti;
2. il potenziamento della semantica di contenuti digitali delle istituzioni culturali e delle pubbliche amministrazioni tramite collegamenti;
3. la creazione di valore culturale ed economico, a partire dal potenziamento della collaborazione tra i settori pubblico e privato e, quindi, dall'interoperabilità dei loro contenuti (nel rispetto delle licenze) pubblicati nel web.

La LOD cloud espone dataset di natura differente, divisi in 7 categorie:

- Cultura e media;
- Geografia;
- Pubblicazioni;
- Contenuti degli utenti;
- Governo;
- Contenuti interdisciplinari;
- Scienze della vita.

Per ottenere la pubblicazione dei propri dati sulla nuvola occorre rispettare i requisiti dichiarati dai curatori del progetto (<http://lod-cloud.net/>), producendo il proprio dataset in conformità coi principi dei linked data:

- uso di <http://> oppure di <https://>URI;
- strutturazione dei dati in formato RDF (RDFa, RDF/XML, Turtle, N-Tuple);
- ricchezza del dataset (deve contenere almeno 1000 triple);
- collegamento del dataset tramite link RDF a dataset già presenti nel diagramma; il dataset deve contenere URI che puntino a dataset esterni; occorre che il dataset contenga almeno 50 collegamenti (si tratta di un numero convenzionale per garantire un numero minimo di collegamenti);
- possibilità d'accesso all'intero dataset tramite RDF crawling, dump RDF o SPARQL endpoint.

²² <<https://openlibrary.org/>>.

²³ <<http://lov.okfn.org/dataset/lov/>>.

6. LICENZE DI PUBBLICAZIONE DEI DATI

Le licenze rappresentano un aspetto cruciale per l'uso dei dati. Nel contesto degli open data, le licenze più diffuse sono basate su *Creative Commons* (CC),¹ ormai standard de facto a livello internazionale, prodotto da un network di esperti articolato in oltre 70 gruppi di lavoro (*affiliate networks*²) che, nelle rispettive giurisdizioni nazionali, curano la promozione, l'aggiornamento, l'adattamento e la traduzione di un set di licenze. Ciascuna licenza CC³ è espressa in tre diversi formati:

- una sintesi destinata alle persone comuni (*common deed*);
- una versione estesa in linguaggio tecnico-giuridico (*legal code*);
- i metadati leggibili solo dai computer.

Le licenze CC sono piuttosto semplici da applicare per chi crei e pubblica dati nel web, tramite l'utilizzo di un tool, con interfaccia per semplificarne l'uso, che produce la licenza sulla base delle domande poste a chi lo utilizzi, del tipo: “Vuoi che la risorsa sia modificabile?”; “È possibile che la risorsa sia usata per scopi commerciali?”.⁴ Sulla base delle risposte, il sistema produce una licenza che viene associata alla risorsa.

Licenza Creative Commons Zero (CC0): consente qualsiasi utilizzazione, comprese quelle per finalità commerciali. Associando a una risorsa la dichiarazione CC0 si afferma che essa è di pubblico dominio o che il titolare rinuncia definitivamente a tutti i diritti di privativa nel modo più ampio possibile consentito dalle norme di protezione del diritto d'autore.

Licenza Creative Commons Attribuzione (CC-BY): permette al soggetto utilizzatore di riprodurre, distribuire, comunicare, esporre, rappresentare, nonché modificare e usare un insieme di dati a fini scientifici e commerciali con l'unico obbligo di rico-

¹ <<http://creativecommons.org/>>.

² <https://wiki.creativecommons.org/CC_Affiliate_Network>.

³ <<http://creativecommons.org/licenses/>>. Le versioni italiane sono all'URL <<http://www.creativecommons.it/Licenze>>.

⁴ <<http://creativecommons.org/choose/>>.

noscere la titolarità dell'opera. Per l'attribuzione, caratteristica delle licenze CC-BY, il licenziatario deve provvedere alla citazione di:

- a. autore originale o titolare dei diritti;
- b. terze parti designate, se esistenti;
- c. titolo del documento;
- d. Uniform Resource Identifier (URI) che il licenziante specifichi dover essere associato con il documento;
- e. nel caso di documenti rielaborati o di opere derivate, l'attribuzione dovrà essere esplicita, così da non ingenerare confusione rispetto all'origine del documento.

La licenza CC-BY è stata rilasciata nel 2013 ed è espressione di una nuova generazione di licenze aperte.⁵ Le licenze preesistenti di tipo CC-BY- includono ulteriori limitazioni introdotte dalle seguenti specifiche o attributi:

- *Share Alike* (SA): l'opera può essere riutilizzata solo allo stesso modo dell'originale, ossia con la stessa licenza d'uso;
- *Non Commercial* (NC): vieta il riuso per finalità commerciali;
- *No Derivative Works* (ND): vieta la creazione di opere derivate.

La combinazione di questi attributi genera le seguenti licenze:

- Attribution Non-commercial No Derivatives (by-nc-nd)
- Attribution Non-commercial Share Alike (by-nc-sa)
- Attribution Non-commercial (by-nc)
- Attribution No Derivatives (by-nd)
- Attribution Share Alike (by-sa)

Le licenze CC considerate compatibili con la *open definition*⁶ adottata dalla Open Knowledge Foundation (OKFn)⁷ sono CC0, CC-BY e CC-BY-SA.

In Italia, l'art. 2 del D. Lgs. n. 36/2006 fornisce la seguente

⁵ L'annuncio è del 25 novembre 2013, <<http://creativecommons.org/weblog/entry/40768>>.

⁶ <<http://opendefinition.org/>>.

⁷ <<https://okfn.org/opendata/>>.

definizione di licenza di riuso dei dati: “Il contratto, o altro strumento negoziale, redatto ove possibile in forma elettronica, nel quale sono definite le modalità di riutilizzo dei documenti delle pubbliche amministrazioni o degli organismi di diritto pubblico”. FormezPA⁸ ha sviluppato apposite licenze aperte *Italian Open Data License* (IODL) a uso delle pubbliche amministrazioni:

- IODL 2.0:⁹ permette di consultare, estrarre, copiare e pubblicare i dati liberamente, anche a scopo commerciale, a condizione di citare la fonte; offre, inoltre, la possibilità di creare un’opera derivata integrando dataset diversi;
- IODL 1.0,¹⁰ simile alla prima, ma con l’obbligo di pubblicare o condividere le opere derivate con la stessa licenza o con una licenza compatibile.

Tabella di comparazione delle licenze CC e IODL:

	CC0	CC BY	CC BY SA	CC BY SA NC	IODL 1.0	IODL 2.0
Portabilità interlingua	X	X	X	X		
Riconoscibilità internazionale logo	X	X	X	X		
Uso per fini commerciali	X	X	X		X	X
Uso gratuito	X	X	X	X	X	X
Possibilità di modificare i dati	X	X	X	X	X	X
Possibilità di preservare la paternità		X	X	X	X	X
Compatibilità inter-licenza		X	Y ¹¹		X	X

⁸ <<http://www.formez.it/>>.

⁹ <<http://www.dati.gov.it/iodl/2.0/>>.

¹⁰ <<http://www.formez.it/iodl/>>.

¹¹ Compatibile solo con IODL 1.0.

Altre licenze aperte, diffuse a livello internazionale, compatibili con la definizione della OKFn sono:

- *Open Data Commons Public Domain Dedication and License* (ODC PDDL): applicabile ai database e ai dati in essi presenti, pone in pubblico dominio sia l'uno che gli altri;¹²
- *Open Data Commons Attribution* (ODC-BY): molto simile a CC-BY, consente qualsiasi utilizzazione a condizione di citare la fonte;
- *Open Database License* (ODbL):¹³ più limitata della precedente, permette di copiare, condividere, distribuire, rielaborare e produrre opere derivate dal database, a condizione di citare la fonte, e di condividere allo stesso modo, con licenza aperta, i prodotti estratti e rielaborati;
- *GNU Free Documentation License* (GNU FDL):¹⁴ applicabile a manuali e altri tipi di documentazione, consente la riproduzione, la ripubblicazione, la rielaborazione di opere per finalità commerciali e non commerciali;
- *GNU General Public License* (GNU GPL):¹⁵ applicabile a software e ad altre creazioni, ne permette la più ampia e libera riutilizzazione a condizione di citare la fonte e condividere allo stesso modo quanto realizzato.

Licenze BDS:¹⁶ applicabili al software, sono compatibili con la GNU GPL.

¹² <<http://opendatacommons.org/licenses/pddl/1-0/>>.

¹³ <<http://opendatacommons.org/licenses/odbl/1.0/>>.

¹⁴ <<http://www.gnu.org/copyleft/fdl.html>>

¹⁵ <<http://www.gnu.org/copyleft/gpl.html>>.

¹⁶ <http://it.wikipedia.org/wiki/Licenze_BSD>.

7. NORMATIVA ITALIANA E DISPOSIZIONI EUROPEE

La normativa sui linked data richiederebbe una trattazione specifica. Riportiamo qui i riferimenti principali alla normativa italiana formulata nelle *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico* del 2014 (paragrafo 3.1), allegate alla Determinazione Commissariale n. 95/2014 del 26-06-2014, emanate dall'Agenzia per l'Italia digitale della Presidenza del Consiglio dei ministri;¹ riportiamo, inoltre, riferimenti a disposizioni dell'Unione europea.

7.1 Normativa italiana

“In ambito nazionale, la nuova normativa per i dati di tipo aperto e le norme, ancor più recenti, in materia di trasparenza dell'azione amministrativa e di pubblicazione tracciano un quadro prescrittivo per le amministrazioni chiaramente orientato alla pubblicazione di dati in formato di tipo aperto e liberamente utilizzabili, salvo casi eccezionali da individuare sulla base di criteri oggettivi e trasparenti, come in materia di protezione dei dati personali.

In particolare, l'art. 9 del DL n. 179/2012, convertito in Legge n. 221/2012, ha interamente riscritto l'art. 52 del CAD [Codice dell'Amministrazione digitale] sull'accesso telematico e riutilizzo dei dati delle pubbliche amministrazioni. Esso stabilisce che le amministrazioni, i gestori di pubblici servizi e le società partecipate a maggioranza pubblica, inserite nel conto economico consolidato della pubblica amministrazione, ‘pubblicano nel proprio sito web, all'interno della sezione Trasparenza, valutazione e merito (ora Amministrazione trasparente secondo quanto previsto dal D.lgs n. 33 del 14 marzo 2013), il catalogo dei dati, dei metadati e delle relative banche dati in loro possesso ed i regolamenti che ne disciplinano l'esercizio della facoltà di accesso telematico e il riutilizzo, fatti salvi i dati presenti in Anagrafe tributaria’. Lo

¹<http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

stesso articolo stabilisce che ‘i dati e i documenti che le amministrazioni titolari pubblicano senza l’espressa adozione di una licenza si intendono rilasciati come dati di tipo aperto’ (principio dell’Open Data by default). L’eventuale adozione di una licenza è motivata ai sensi delle linee guida nazionali”.

Particolarmente innovativa è la previsione, sempre definita all’art. 9 del DL suddetto, relativa alle clausole da inserire nei capitolati e negli schemi di contratto per i servizi ICT (Information and Communication Technology), che prevedono la raccolta e la gestione di dati pubblici. Le pubbliche amministrazioni devono prevedere “clausole idonee a consentire l’accesso telematico e il riutilizzo, da parte di persone fisiche e giuridiche, di tali dati, dei metadati, degli schemi delle strutture di dati e delle relative banche dati”. La norma stabilisce inoltre che “le attività volte a garantire l’accesso telematico e il riutilizzo dei dati delle pubbliche amministrazioni rientrano tra i parametri di valutazione della performance dirigenziale”. Infine, sempre l’art. 9 riscrive l’art. 68, comma 3, del CAD sulla definizione dei dati di tipo aperto identificandone le principali caratteristiche, come riportato nella sezione 3.3, e pone in capo all’Agenzia per l’Italia digitale il compito di individuare i casi eccezionali di dati aperti rilasciati a tariffe superiori ai costi marginali per la loro riproduzione e divulgazione. Se per i dati pubblici il CAD e le nuove opportunità aperte dagli open data hanno portato alla formulazione di nuove regole d’accesso e diffusione, in ambito statistico l’attenzione alla fornitura, diffusione e accesso alle informazioni da parte degli utenti risulta antecedente rispetto alla nascita del paradigma dell’open data, pur se in origine con specifico riferimento al concetto di dato aggregato.

Nell’ambito dei dati statistici esistono normative di riferimento quali, per esempio, il D.lgs. n. 322 del 6 settembre 1989 – *Norme sul Sistema statistico nazionale e sulla riorganizzazione dell’Istituto nazionale di statistica* – nel quale viene evidenziata e ripresa la necessità di interconnettere i sistemi informativi delle pubbliche amministrazioni e degli enti che fanno parte del Sistema statistico nazionale, garantendo l’accesso alle informazioni prodotte dalle amministrazioni. Il riferimento è presente nell’art. 6, del D.lgs. n. 322/1989, che disciplina i compiti degli uffici di statistica; “essi:

- a) promuovono e realizzano la rilevazione, l’elaborazione, la

diffusione e l'archiviazione dei dati statistici che interessano l'amministrazione di appartenenza, nell'ambito del programma statistico nazionale;

- b) forniscono al Sistema statistico nazionale i dati informativi previsti dal programma statistico nazionale relativi all'amministrazione di appartenenza, anche in forma individuale ma non nominativa ai fini della successiva elaborazione statistica;
- c) collaborano con le altre amministrazioni per l'esecuzione delle rilevazioni previste dal programma statistico nazionale;
- d) contribuiscono alla promozione e allo sviluppo informatico a fini statistici degli archivi gestionali e delle raccolte di dati amministrativi".

Per i compiti suddetti, "gli uffici di statistica hanno accesso a tutti i dati statistici in possesso dell'amministrazione di appartenenza, salvo eccezioni relative a categorie di dati di particolare riservatezza espressamente previste dalla legge. Essi possono richiedere all'amministrazione di appartenenza elaborazioni di dati necessarie alle esigenze statistiche previste dal programma statistico nazionale". Inoltre, in base all'art. 10 che regola l'accesso ai dati statistici, il decreto recita: "I dati elaborati nell'ambito delle rilevazioni statistiche comprese nel programma statistico nazionale sono patrimonio della collettività e vengono distribuiti per fini di studio e di ricerca a coloro che li richiedono fermi restando i divieti di cui all'art. 9"; esso prevede la tutela del segreto statistico, di conseguenza, la protezione dei dati personali, consentendo la comunicazione e la diffusione unicamente di dati secondo modalità che non rendano identificabili gli interessati.

7.2 Disposizioni europee

L'Agenda digitale europea fornisce indicazioni precise: la prima è l'azione n. 3 relativa allo sviluppo del mercato unico digitale; essa individua alcune linee strategiche per gli Stati membri, relativamente all'apertura di dati per il loro riutilizzo. La seconda indicazione è l'azione n. 85 relativa alla revisione dell'accesso pubblico alla direttiva sulle informazioni ambientali.

Altrettanto importanti sono le disposizioni contenute nella di-

rettiva 2013/37/UE che modifica la direttiva 2003/98/CE sull'informazione del settore pubblico (PSI). Recepita in Italia dal D.lgs n. 36 del 2006 pubblicato sulla "Gazzetta ufficiale" del 26 giugno 2013, la nuova direttiva ha ampliato gli obblighi degli enti pubblici di rendere liberamente riutilizzabili i dati in loro possesso, a scopo commerciale e non commerciale, a condizione che le informazioni non siano escluse dal diritto d'accesso, in conformità alla normativa sulla protezione dei dati; ha, inoltre, esteso l'ambito d'applicazione degli obblighi alle istituzioni culturali (biblioteche, musei e archivi), in precedenza escluse. Sono consentite tariffazioni per coprire i costi di riproduzione, fornitura e diffusione; gli istituti culturali possono concedere i diritti esclusivi di utilizzazione solo per finalità specifiche, come, per esempio, la copertura dei costi di digitalizzazione, ma per una durata limitata nel tempo. La regola generale è che le amministrazioni rendano disponibili i loro dati e documenti, ove possibile, in formati aperti e "machine-readable", accompagnandoli dai relativi metadati; i formati e i metadati dovrebbero essere conformi a formati aperti standard; le licenze da associare ai dati non dovrebbero limitare le possibilità di riutilizzo dei dati e documenti.

Il G8 tenuto a Londra nel giugno 2013 è intervenuto su tre direttrici:

1. la stesura di una *Carta degli open data* che sancisca principi basilari che gli Stati dovrebbero recepire per rendere disponibili e riutilizzabili i dati raccolti e gestiti dalle pubbliche amministrazioni;
2. l'impegno da parte degli Stati di far propri i principi della *Carta*;
3. un piano d'azione per il rilascio di talune tipologie di dati.

Nella Comunicazione 882/2011 l'Unione europea ha precisato che gli investimenti sugli open data rispetteranno appieno il copyright.²

Alla luce di queste direttrici, emerge che l'Agenda nazionale e le *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico* del 2014 sono allineate alle disposizioni internazionali.

² <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>>.

8. BIBLIOTECHE E WEB SEMANTICO

Se si volesse rappresentare con pochi aggettivi il mondo delle biblioteche, degli archivi e dei musei nel ruolo di *distributori dell'informazione* bisognerebbe pensare a una combinazione di attributi, tra cui:

- *esteso*: si pensi a quanti enti culturali sono presenti nel territorio nazionale e internazionale; il concetto dello spazio e delle relative distanze, che un tempo costituiva un elemento importante nella valutazione della disponibilità delle risorse informative, oggi ha un'altra dimensione, superato dalla realtà del web e dalla possibilità di pensare a un universo globale in cui sia possibile operare e vivere;
- *eterogeneo*: le biblioteche, i musei e gli archivi sono differenti per natura, grandezza e finalità; rappresentano realtà diverse in termini di tipologia, raccolta, dimensione e tradizione. Nell'eterogeneità degli istituti della memoria registrata includiamo livelli molto distanti di informatizzazione e di capacità di offerta di servizi, spesso dipendenti da fattori economici, politici e organizzativi;
- *disseminato*: gli istituti sono spesso distribuiti sul territorio senza forme di contatto e collaborazione; i progetti di cooperazione e di realizzazione di rete hanno una storia e un radicamento completamente differenti da Paese a Paese, da città a città, da comunità a comunità, tant'è che difficilmente si può pensare a un modello omogeneo di sistema integrato; la disseminazione dell'informazione è una tematica aperta;
- *ricco*: il patrimonio informativo offerto dagli istituti culturali è cospicuo e in molti casi unico.

Una caratteristica accomuna e caratterizza la tradizione degli istituti della memoria: l'attenzione radicata a strutturare il dato e a proporre studi e standard finalizzati alla registrazione dell'informazione (o, in linguaggio contemporaneo, del metadata). Per questa consolidata sensibilità, il patrimonio informativo *prodotto* dalle biblioteche, e dunque i cataloghi e i repertori di

controllo (soggettari, authority file, etc.), più che i documenti a cui questi rimandano, hanno attirato su di sé l'attenzione del mondo del web semantico, che di questi silos informativi ha individuato il potenziale enorme di ricchezza e riuso. La novità dell'ultimo decennio è, però, il fenomeno inverso: la curiosità del mondo delle biblioteche per il web semantico, la consapevolezza sempre più ampia di vantaggi e opportunità reciproche in questa condivisione.¹ Eppure, un occhio poco attento o ancora poco esperto potrebbe non valutare la portata innovativa e rivoluzionaria dell'interesse reciproco e del nuovo corso che le biblioteche potrebbero avviare, facilitando e accelerando l'incontro. Cerchiamo di capire le ragioni di una collaborazione necessaria e inevitabile.

8.1 I dati delle biblioteche nel web semantico

La tradizione di pubblicare i dati del catalogo nel web è radicata nella storia delle biblioteche come in poche altre comunità. Il formato MARC è nato proprio per lo scambio dei metadati catalografici tra istituzioni diverse; il nome esteso – MACHine Readable Cataloging – indica la caratteristica del formato di essere letto e interpretato dalle macchine; ciò non deve, tuttavia, indurre nell'errore di credere che le biblioteche, oggi, siano conformi a quanto richiesto dagli attuali paradigmi del web, affinché i dati del proprio catalogo siano *del web* e non più solo *nel web*, e che i record bibliografici e d'autorità siano fruibili da parte di agenti elettronici. Esiste, infatti, una distanza da colmare tra la realtà dei dati prodotti dagli istituti della memoria registrata e le tecniche per la realizzazione del web semantico; occorre che le biblioteche compiano un adeguamento tecnologico e soprattutto di mentalità. La necessità del cambiamento è stata formulata dalle biblioteche in varie forme e occasioni. La più significativa nella direzione del web semantico è *On the record*, il report della Library of Congress Working Group on

¹ La creazione delle prime ontologie legate specificatamente al mondo delle biblioteche è del 2005; tra queste: FRBR Core, FRBR Extended e SKOS.

the Future of Bibliographic Control,² reso pubblico il 9 gennaio 2008, in cui si richiama alla necessità di:

- trasformare la descrizione testuale in set di dati usabili per processi ed elaborazioni automatiche da parte di macchine;
- rendere gli elementi di dati univocamente identificabili all'interno del contesto informativo del web;
- assicurare la compatibilità dei dati con le tecnologie e gli standard del web;
- usare un linguaggio trasversale e interoperabile nella realtà del web.

La dichiarazione di orientamento verso il web semantico è chiara e ufficializzata, confermando quanto anticipato con la pubblicazione di schemi dati e ontologie in RDF, già dal 2005.

8.2 Dal metadato al *dato significativo*

Da dove parte l'analisi sul record bibliografico e perché esso, così strutturato e storicamente formalizzato, si è rivelato inadeguato nel contesto del web semantico? La storia dei cataloghi delle biblioteche e degli archivi mostra un impiego antico e diffuso dei metadati, intesi come *informazioni vicarie* della risorsa. Nell'accezione tradizionale del termine, un metadato è un dato su un altro dato: nel record descrittivo di un libro o di un film, per esempio, i metadati sono il titolo, il nome dell'autore o del regista, l'anno di pubblicazione. Le caratteristiche principali dei metadati nell'accezione tradizionale sono:

- la natura artefatta, costruita sulla risorsa;
- la finalità descrittiva;
- l'usabilità da parte di macchine; il metadato dev'essere strutturato in modo che sia processabile e, dunque, utilizzabile da un computer.

I campi del record utilizzati per esprimere il concetto di metadato rispondono in modo completo alle prime due caratteri-

² <www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.

stiche: sono costruzioni aggiunte all'oggetto, che è la risorsa, e hanno la finalità di descriverla, cioè di fornirne i dati distintivi.

Nel web semantico le due caratteristiche non sono sufficienti a rendere adeguato questo metadato, se non accompagnate dalla terza caratteristica che, però, dev'essere estesa nel significato, abbracciando l'accezione di *comprensibilità da parte delle macchine*. Il concetto di metadato, nel contesto del web semantico, estende, pertanto, il suo significato e diventa sinonimo di *dato processabile e interpretabile da una macchina*: le strutture sintattiche del metadato, creano, così quello strato significativo del web che rende espliciti ad agenti elettronici i significati delle risorse che descrivono e delle loro relazioni con altre risorse.

8.3 Record bibliografico vs RDF

La registrazione catalografica, intesa come un insieme di metadati descrittivi della risorsa e di meccanismi per garantirne l'accesso, è sempre stata al centro dell'attenzione e delle attività delle biblioteche. Il passaggio dal catalogo cartaceo al catalogo elettronico, che ha avuto la massima espressione con la nascita del formato MARC a metà anni Sessanta, con il relativo accrescimento dell'interesse verso gli standard e le formattazioni dell'informazione (per esempio, con l'ISBD, dalla fine del medesimo decennio), ha amplificato la centralità del record, che è diventato il catalizzatore di una serie di servizi differenti offerti a utenti di vario tipo: dai programmi di catalogazione partecipata, ai processi di interlibrary loan (prestito interbibliotecario), alla nascita delle reti cooperative. La centralità del record ha, tuttavia, comportato una serie di inconvenienti:

- la produzione di versioni varianti di record standard, per rispondere a esigenze locali, non sempre giustificate, di singole biblioteche;
- la redazione di record molto dettagliati, con elevato costo di produzione, ma anche di alta qualità;
- la creazione di *silos chiusi* e per lo più non usabili al di fuori del mondo delle biblioteche;
- l'annullamento degli sforzi compiuti per creare i record, poiché non trasferibili nello spazio del web, soprattutto del web semantico.

Nel web semantico il record MARC è considerato una fonte d'informazione preziosa per l'enorme quantità di *dati* che contiene e per l'elevato grado di *semantica* in esso presente: il set di metadati raccolto per descrivere la manifestazione *As you like it* di Shakespeare, nell'esempio riportato sotto, è ricchissimo e ogni tag rimanda a uno specifico *significato* del campo.

Esempio:

LC Control No.:	2009025024
LCCN	
Permalink:	http://lcn.loc.gov/2009025024
000	02384cam a2200445 a 450
001	15789102
005	20100204124659.0
008	090624s2009 enka b 000 0 eng
020	<u>1A</u> 9780521519748 (hbk.)
020	<u>1A</u> 0521519748 (hbk.)
020	<u>1A</u> 9780521732505 (pbk.)
020	<u>1A</u> 0521732506 (pbk.)
035	<u>1A</u> (OCoLC)ocn416716652
040	<u>1A</u> DLC <u>1C</u> DLC <u>1D</u> YDXCP <u>1D</u> BWKUK <u>1D</u> BWK <u>1D</u> FDA <u>1D</u> UKM <u>1D</u> BTCTA <u>1D</u> CEF <u>1D</u> CDX <u>1D</u> DLC
050	00 <u>1A</u> PR2803.A2 <u>1B</u> H35 2009
082	00 <u>1A</u> 822.3/3 <u>12</u> 22
100	1 <u>1A</u> Shakespeare, William, <u>1D</u> 1564-1616.
245	10 <u>1A</u> As you like it / <u>1C</u> edited by Michael Hattaway.
250	<u>1A</u> Updated ed.
260	<u>1A</u> Cambridge, UK ; <u>1A</u> New York : <u>1B</u> Cambridge University Press, <u>1C</u> 2009.
300	<u>1A</u> xv, 240 p. : <u>1B</u> ill. ; <u>1C</u> 24 cm.
490	1 <u>1A</u> New Cambridge Shakespeare
504	<u>1A</u> Includes bibliographical references (p. 240).
505	0 <u>1A</u> Introduction. Journeys ; Plays within the play ; Theatrical genres ; Pastoral ; Counter-pastoral ; The condition of the country ; Politics ; 'Between you and the women the play must please' ; Gender ; Nuptials ; Sources ; Date and occasion ; Stage history ; Recent critical and stage interpretations -- Note on the text -- List of characters -- The play -- Textual analysis -- Appendixes : 1. An early court performance? ; 2. Extracts from Shakespeare's principal source, Lodge's <i>Rosalind</i> ; 3. The songs.
650	0 <u>1A</u> Fathers and daughters <u>1V</u> Drama.
650	0 <u>1A</u> Exiles <u>1V</u> Drama.
600	10 <u>1A</u> Shakespeare, William, <u>1D</u> 1564-1616. <u>1T</u> As you like it.
655	7 <u>1A</u> Pastoraldrama. <u>12</u> gsafd
655	7 <u>1A</u> Comedies. <u>12</u> gsafd
700	1 <u>1A</u> Hattaway, Michael.

La semantica espressa nel record è comprensibile solo all'occhio umano. I *singoli elementi informativi* (vedi sotto) perdono di significato se isolati dal record:

- xv, 240 p.
- Includes bibliographical references (p. 240).
- Cambridge (UK)
- New York
- Cambridge University Press, 2009.
- edited by Michael Hattaway.
- Updated ed.
- Fathers and daughters- Drama.

Trasferito nel web tradizionale, gran parte del *significato* del record in formato MARC viene perduto; molte informazioni sono trattate dai motori di ricerca come *blocchi di testo* e, dunque, non comprese singolarmente e inequivocabilmente. L'intero record d'altra parte, con la sua complessità e ricchezza, diventa riusabile solo da chi trasferisce sul proprio sistema una copia del medesimo record per descrivere la medesima manifestazione; nel caso, per esempio, di una nuova edizione della medesima opera, la normativa catalografica e la tecnologia informatica richiedono la creazione di un nuovo record, che potrebbe differire dal precedente solo per il campo edizione e data. Da ciò possiamo immaginare la moltiplicazione smisurata dei record all'interno di ciascun catalogo. Inoltre i record pubblicati tramite gli OPAC in formato MARC sono stati usati pressoché esclusivamente nell'ambito delle biblioteche; la comunità degli editori, per esempio, non lo utilizza per le procedure di scambio dati con le biblioteche o i fornitori di servizi affini.

Gli *elementi singoli* possono riacquistare nel web il significato che l'intero record conferiva loro se i blocchi monolitici di informazioni (record) sono trasformati in elementi o dati significativi; in ciò consiste il passaggio dal web tradizionale al web semantico: gli stessi metadati codificati nei tag del record, tradotti in *asserzioni*, riconquistano il *valore semantico* originario, che diventa *comprensibile e usabile da una macchina*.

I dati acquistano un significato in sé valido, componendosi, tramite il processo di mappatura e conversione, in asserzioni (o triple auto-esplanti), riusabili da comunità diverse. Se il record

bibliografico poteva essere utilizzato in un'altra biblioteca solo nel caso in cui la risorsa descritta fosse la stessa, il singolo dato può, invece, essere riusabile in situazioni differenti e in descrizioni diverse, essendo identificato e definito nel suo valore distintivo, come dato autoconsistente, nell'universo del web:

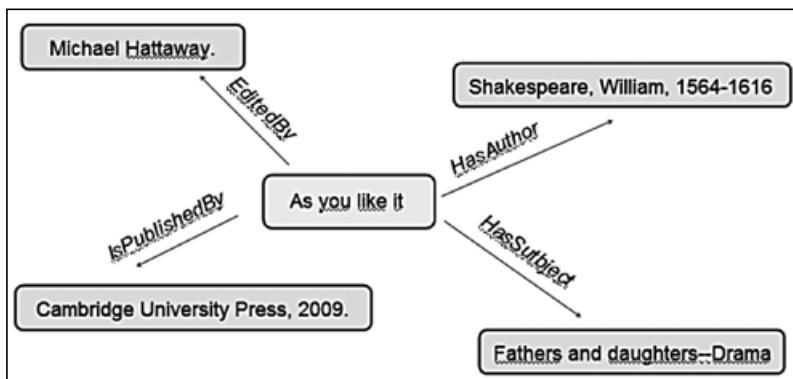


Figura 34-a – Traduzione dei dati del record bibliografico in asserzioni.

Ciascun elemento dell'asserzione così prodotta può essere registrato con un URI, che *collega* il dato ad altri presenti sul medesimo dataset o su dataset esterni. Il trasferimento di queste asserzioni comprensibili alle macchine nel web trasforma i dati contenuti negli originali record MARC in *parti integranti del web semantico*.

8.4 Nuovi modelli catalografici e linked data

L'adozione delle logiche e delle tecnologie del web semantico può avere un impatto differente nei vari contesti di adozione: pubblicare open data o linked open data in un contesto industriale, per esempio, può significare registrare i propri dati su tabelle Excel o convertire dei database in RDF, senza probabilmente cambiare i processi produttivi del dato stesso. Nel caso delle biblioteche (e degli archivi prevalentemente in ambito anglosassone) il passaggio ha generato un ripensamento profondo del modello del catalogo finora conosciuto, producendo la rifondazione dei principi, degli standard delle normative ca-

talografiche. I *Principi internazionali di catalogazione* (ICP)³ del 2009 emanati dall'IFLA parlano di dati e non più di record.

Roy Tennant, nel 2002, in un articolo pubblicato su "Library journal" dal titolo significativo *MARC must die*,⁴ sancisce l'inizio di un nuovo corso evolutivo, che si conclude con la dichiarazione ufficiale della morte di quel formato: la pagina wiki del MARC è stata significativamente (e autoironicamente) rinominata in *MARC is dead*.⁵

RDA (*Resource Description and Access*), concepito inizialmente come un'evoluzione di AACR2 (*Anglo-American Cataloguing Rules*), apre nuove prospettive presentandosi come standard internazionale per la metadatazione delle risorse nell'era digitale; le linee guida (piuttosto che un codice di catalogazione) si muovono all'interno della filosofia del web semantico e dei linked data; favoriscono, per esempio, la pubblicazione di ontologie e vocabolari controllati in RDF, così come dedicano molto spazio alla costruzione di dati di qualità (record d'autorità) che possano essere riutilizzati in contesti diversi e per molteplici scopi. Il modello di FRBR, che si esprime in *entità – relazioni – attributi* (gli *elementi* di RDA) è proposto nelle linee guida come uno schema dati perfetto per l'evoluzione delle norme e degli standard in termini di linked data.

Lo schema della figura 34 mostra le entità, con alcune relazioni e attributi possibili: l'immagine esprime il modello a grafo delle risorse nel web semantico, con una particolare attenzione alle relazioni esplicite tra le risorse.⁶ La tappa più significativa e di maggior rilievo per la comunità catalografica e per il futuro del

³ IFLA. *IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its glossary*, by Barbara Tillett & Ana Lupe Cristán (eds.). München: Saur, 2009, <<http://www.ifla.org/publications/ifla-series-on-bibliographic-control-37>>; traduzione italiana: *Dichiarazione di Principi internazionali di catalogazione*, curata dall'ICCU, <www.ifla.org/files/assets/cataloguing/icp/icp_2009-it.pdf>; per un commento, Mauro Guerrini, con Giuliano Genetasio, *I Principi internazionali di catalogazione (ICP). Universo bibliografico e teoria catalografica all'inizio del 21. secolo*. Milano: Editrice Bibliografica, 2012.

⁴ <<http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>>.

⁵ <http://marc-must-die.info/index.php/Main_Page>.

⁶ Parte dello schema è stato utilizzato da Barbara B. Tillett nel semi-

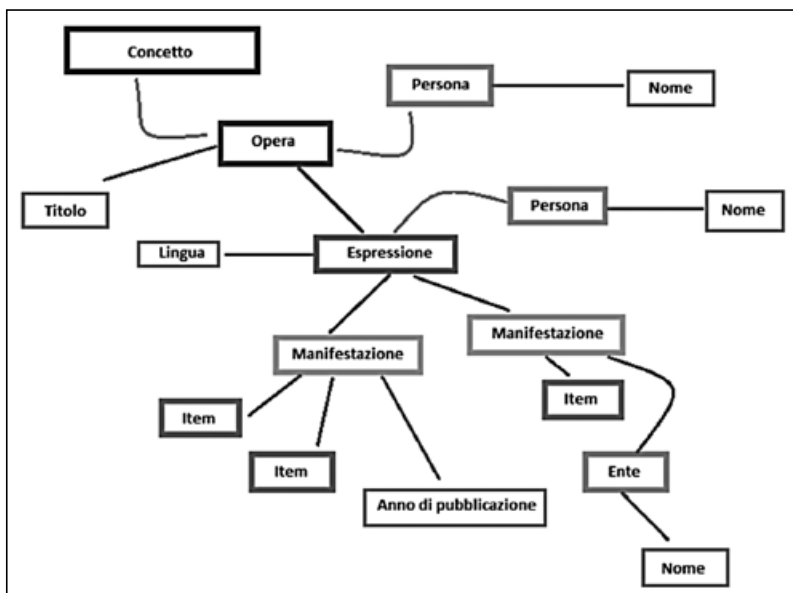


Figura 34-b – Modello di FRBR proposto in RDA come schema per l’evoluzione delle norme in termini di linked data

controllo bibliografico universale è stata la riflessione partita dalla sperimentazione di RDA dal 2010: il risultato più consistente è la consapevolezza che il formato piatto del record MARC21 è inadeguato a esprimere le *relazioni* tra entità bibliografiche che il modello FRBR e lo standard RDA considerano fondamentali. A partire da questa conclusione, la Library of Congress ha annunciato nel maggio 2011, la *Bibliographic Framework Transition Initiative*, dichiarando, nell’ottobre 2011: “The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model”.⁷

La transizione dal web tradizionale al web semantico con l’adozione ufficiale della tecnologia a esso collegata è così sancita

nario su RDA organizzato dal Master biennale in Catalogazione dell’Università di Firenze, con Casalini Libri, ICCU (Linked Heritage) e AIB Toscana, Firenze, 18-20 marzo 2013.

⁷ <<http://www.loc.gov/marc/transition/news/framework-103111.html>>.

ta. Una testimonianza interessante della transizione è la nascita del W3C Library Linked Data Incubator Group, il cui scopo è sostenere e valorizzare l'interoperabilità globale dei dati delle biblioteche nel web, favorendo il confronto tra esperti del web semantico.⁸ Con il report *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services* (BIBFRAME),⁹ pubblicato dalla Library of Congress il 21 novembre 2012, la prospettiva è tracciata.¹⁰

8.5 BIBFRAME. Verso un ecosistema bibliografico

BIBFRAME è il risultato di un percorso di riflessione sulle funzioni del catalogo, sull'attualità del MARC e delle norme catalografiche e sulle nuove tipologie di risorsa presenti nell'universo bibliografico; esso rappresenta il primo passo nella definizione di una road map che dovrebbe proporre un nuovo ambiente bibliografico, che prevede l'interconnessione di entità generate da fonti differenti e l'enunciazione di strategie e strumenti capaci di supportare l'evoluzione in atto nel circuito della comunicazione globale; dovrebbe, in particolare, portare all'enunciazione di un nuovo *ecosistema bibliografico*.¹¹ Le nuove istruzioni per la descrizione e l'accesso alle risorse dovrebbero caratterizzarsi per:

- un alto livello di analisi e *identificazione* del dato;
- un'enfaticizzazione delle *relazioni*;
- un impiego diffuso dei vocabolari controllati;

⁸ W3C Incubator Group, *Library Linked Data Incubator Group Final Report*, October 25, 2011, <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025>>.

⁹ <www.loc.gov/marc/transition/pdf/marclld-report-11-21-2012.pdf>.

¹⁰ Mauro Guerrini, *BIBFRAME. Per un nuovo ruolo delle biblioteche nel contesto del web*, "Digitalia. La rivista del digitale nei beni culturali", vol. 1 (giugno 2014), p. 125-126. <<http://digitalia.sbn.it/article/view/1061/691>>. In allegato la traduzione di BIBFRAME, di Iolanda Cristalli e di Roberto Morellato, disponibile all'indirizzo: <<http://digitalia.iccu.sbn.it>>.

¹¹ Mauro Guerrini, *BIBFRAME. Un'ipotesi di ambiente bibliografico nell'era del web*, in *Il libro al centro. Percorsi fra le discipline del libro in onore di Marco Santoro*. Studi promossi da Rosa Marisa Borraccini, Alberto Petrucciani, Carmela Reale, Paola Zito, a cura di Carmela Reale. Napoli: Liguori, 2014, p. 103-115.

- un accurato dettaglio descrittivo delle risorse;
- un'ampia flessibilità nella gestione delle voci controllate.

Il modello esamina l'evoluzione degli standard e delle norme catalografiche e, dunque, compie una riflessione sul ruolo, la diffusione e le funzioni storiche del MARC21 e sul futuro del controllo bibliografico universale; l'elaborazione teorica si basa sull'esperienza maturata durante fase di test di RDA. La possibile conversione dei dati bibliografici in linked data viene interpretata in termini di evoluzione, piuttosto che di rivoluzione. Le riflessioni elaborate nella fase di sperimentazione di RDA sono approdate a considerazioni di opportunità, se non di necessità, di utilizzo della rete come modello per esprimere e connettere le informazioni, secondo il paradigma del web semantico, ovvero tramite la struttura di URI e tramite la decentralizzazione e atomizzazione del dato. I risultati sperati del processo evolutivo includono la possibilità di giungere a una semplificazione e ottimizzazione dei processi descrittivi delle risorse che permettano, da una parte, la condivisione della descrizione al di fuori della comunità bibliotecaria, dall'altra la possibilità che il lavoro di metadateazione sia distribuito a livello internazionale tra soggetti diversi, così da estendere la copertura del controllo bibliografico e, quindi, di soddisfare al meglio i bisogni degli utenti.

BIBFRAME propone un modello dati estremamente semplice, ispirato a FRBR (ma con differenze significative); distingue tra contenuto concettuale e manifestazione fisica, cioè tra opera e istanza dell'opera; identifica un'entità in maniera non ambigua; evidenzia le relazioni tra le entità, mostrando la natura della correlazione. Il modello è strutturato in 4 classi principali:

- *Opera*: l'essenza concettuale della risorsa catalogata, il cui concetto è assimilabile all'*Opera* e all'*Espressione* di FRBR.
- *Istanza*: una materializzazione individuale dell'opera.
- *Authority*: l'identificazione delle entità che hanno una relazione con l'opera o con l'istanza, per esempio persone, famiglie, enti, concetti, luoghi.
- *Annotazioni*: tutto ciò che correda o arricchisce le risorse del modello BIBFRAME con informazioni aggiuntive, per esempio, dati di esemplare, immagini di copertina, recensioni, TOC, etc.

L'immagine riprodotta di seguito è estratta da BIBFRAME e rappresenta la struttura generale del modello con l'opera al centro del grafo, collegata all'istanza in cui si concretizza, con le relazioni a esse associate.

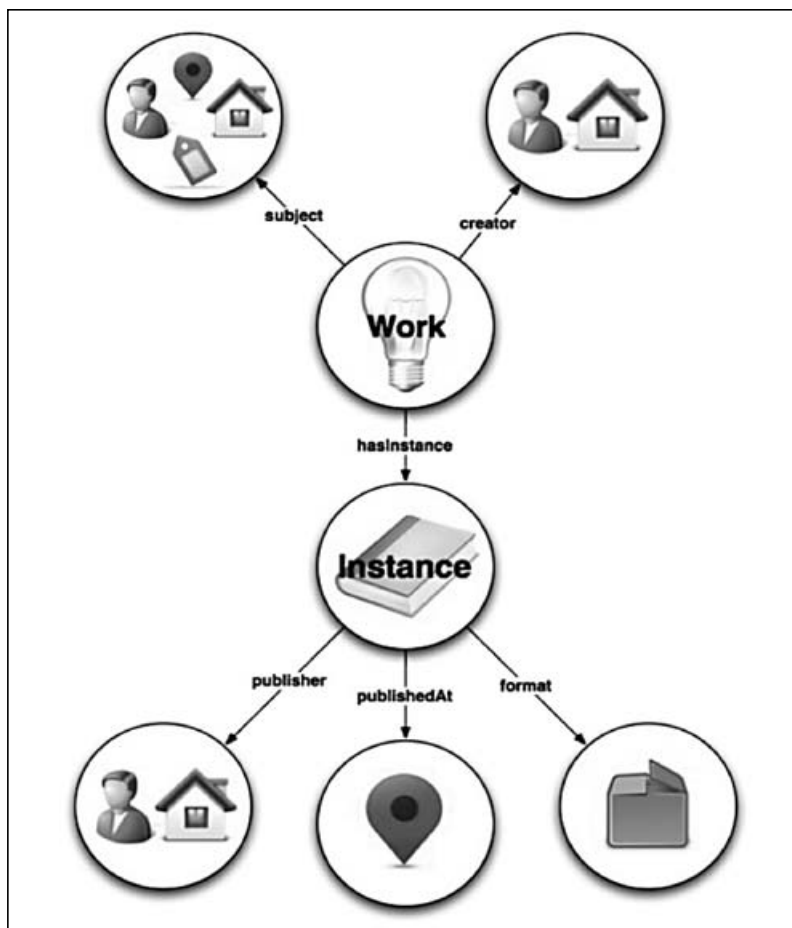


Figura 35 – BIBFRAME: struttura generale del modello

Al centro c'è l'*opera* con link al *creatore* e alla *voce di soggetto*, ovvero al descrittore che esprime il tema di base di cui tratta; l'opera è collegata all'*istanza*, con link all'*editore*, al *luogo* e al *formato* ovvero alla modalità di fruizione: edizione cartacea, edizione online, e-book, etc.

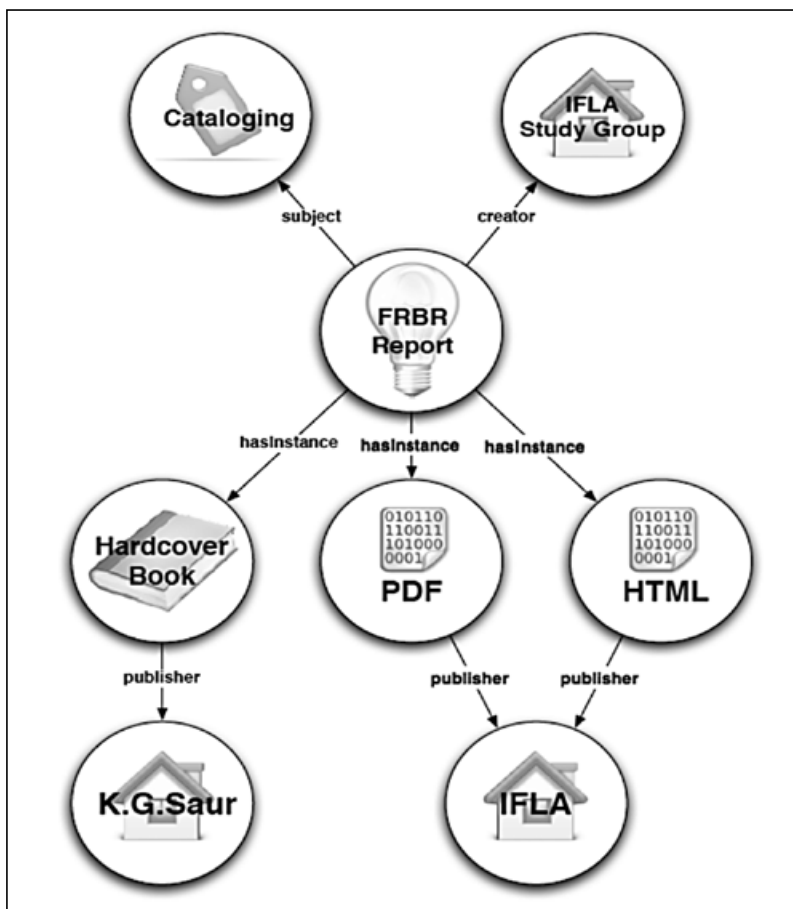


Figura 36 – BIBFRAME: l'esempio dell'opera *FRBR Report*

Il caso concreto in figura 36 presenta al centro l'opera dal titolo *FRBR Report*, con “IFLA Study Group” nel ruolo di ente *creatore* e “Cataloging” come *descrittore di soggetto*. *FRBR Report* ha tre istanze, ovvero tre materializzazioni:

- 1) la versione cartacea con copertina cartonata;
- 2) il pdf;
- 3) la versione online in HTML.

L'edizione cartacea cartonata ha come editore K.G. Saur, mentre il pdf e l'HTML hanno come editore la stessa IFLA.

BIBFRAME indica alla comunità bibliotecaria e ai produttori di software un percorso di cambiamento, con al centro le esigenze informative dell'utente, il riutilizzo dei dati in contesti differenti (interoperabilità), la necessità che le macchine possano essere parte del processo di creazione e di distribuzione dell'informazione. Il progetto prevede di assemblare i dati dell'attuale "risorsa MARC" in una nuova architettura coerente, che consenta la metadatazione cooperativa a un livello descrittivo più granulare.

La traduzione dei record bibliografici in linked data è, pertanto, intesa come base per iniziare un confronto e una discussione all'interno della comunità dell'informazione. Il punto centrale è la convinzione dell'utilizzo del web come luogo per esprimere e connettere informazioni. BIBFRAME, dunque, si propone come modello di riferimento secondo il quale i cataloghi delle biblioteche riusciranno a entrare realmente nel web; un web che sta evolvendosi da una rete di documenti a una rete di dati collegati da relazioni qualificate.

Le tecnologie del web semantico sono già ampiamente diffuse nel mondo dei beni culturali per la pubblicazione dei dati.

Non avviene lo stesso per le modalità e gli strumenti utilizzati per la catalogazione della risorsa, per i quali non abbiamo né realizzazioni, né indicazioni su ciò che questi nuovi sistemi di creazione del dato dovranno garantire. BIBFRAME definisce, tuttavia, alcune linee guida generali, tra cui:

- la prospettiva che gli LMS (Library Management System) di nuova generazione si orientino su una architettura 'a risorsa', dove cioè l'oggetto catalogato sia il singolo elemento o risorsa da trattare, non l'intero record;
- la possibilità che ciascuna risorsa così individuata o creata possa essere collegata ad altre risorse, sul modello entità-relazione di FRBR;
- la possibilità che nell'associazione di attributi a ciascuna risorsa il catalogatore possa attingere da informazioni nel web, esposte secondo i requisiti dei linked data, senza creare ogni volta oggetti nuovi;
- l'impiego di liste controllate di termini e, quindi, l'accesso in fase di costruzione dell'entità ai vocabolari e alle liste di termini pubblicate, per esempio, sull'Open Metadata Registry o su Vocab.org;

- la possibilità di definire in configurazione del sistema la logica di creazione degli URI e di permettere, dunque, che a ogni nuova risorsa creata sia attribuibile automaticamente un URI.

Tutto ciò esprime il quadro generale entro il quale i nuovi LMS vanno ripensati e scritti.

8.6 Integrated Library System (ILS) di prossima generazione

Il passaggio dal record al dato genera una trasformazione profonda negli strumenti di mediazione tra universo bibliografico e utente che nessun attore che operi in questo ambito può ignorare. La nuova generazione di Integrated Library System (ILS) tiene conto delle suggestioni dei linked data, proponendo un nuovo modello di creazione, condivisione e rappresentazione dei dati. Lo sviluppo degli strumenti di elaborazione dei dati sta focalizzandosi su alcuni aspetti (o assunti) fondamentali, in particolare:

- maggiore granularità nella strutturazione dell'informazione;
- utilizzo di vocabolari controllati, di ontologie e di authority file per condividere il più possibile un linguaggio comune e allargare la condivisione alle macchine;
- identificazione chiara degli oggetti e delle entità, tramite codici identificativi univoci e attributi qualificanti;
- creazione, quanto più possibile ampia, di un reticolo di collegamenti tra entità differenti, per fare in modo che ogni oggetto sia connesso ad altri in diverse forme di relazione;
- potenziamento dei processi di cattura e, dunque, di condivisione, dei dati, così da coinvolgere attori diversi, tradizionalmente esclusi da questi processi, come, per esempio, gli editori, che più di altri sono in contatto con gli autori, cioè con coloro che producono informazione (fonti primarie).

Questi principi sono stati assunti come punti di partenza per la formulazione di RDA. La sfida più urgente per i produttori di nuovi sistemi di catalogazione è, perciò, assorbire integralmente i principi dei linked data, semplificandone la realizzazione.

8.6.1 RIMMF (*RDA in Many Metadata Formats*)

Un primo esempio di sistema orientato a RDA è RIMMF (RDA in Many Metadata Formats), sviluppato dalla società The MARC of Quality (TMQ), disponibile al gennaio 2015 in versione beta e scaricabile, per ambiente Windows, con licenze Creative Commons (licenza di uso gratuito, non per uso commerciale). RIMMF è uno strumento illustrativo e didattico, di visualizzazione dei dati, nato per aiutare i catalogatori a pensare in RDA e, dunque, in FRBR, anziché in AACR e MARC. RIMMF non è, quindi, un ILS, né un modulo di catalogazione, perché i dati creati non sono ancora utilizzabili come prodotto, sebbene siano esportabili in vari formati (RDF, XML, MARC). Esso può essere piuttosto considerato un prototipo di come potrebbe essere un'interfaccia impiegando RDA. Il suo valore consiste nell'obiettivo di semplificare il cambiamento di approccio alla catalogazione, che dal processo di produzione del record diventa il processo d'identificazione e metadatazione delle entità FRBR (Opera, Espressione, Manifestazione, Item; Persona, Famiglia, Ente; Concetto, Oggetto, Evento, Luogo). Il cambiamento è evidente nella selezione del modello da utilizzare, in relazione alla risorsa che si sta descrivendo: appare subito chiara l'attenzione posta sulle entità FRBR.



Figura 37 – Selezione del modello di metadatazione secondo le entità FRBR

File Edit View Options Help		
Element Label	Text	RDA Rule
Manifestation		
RDPF Identifier 1	1e100000011	09.rule
Title		2.1
Title Proper	The adventures of Pinocchio	2.1.2
Later Title Proper	Pinocchio	2.1.8
Statement of Responsibility		2.5
Statement of Responsibility Relating to Title Proper	Written by Bonnie Pike, music by M. Michael Fauss	2.4.2
Publication Statement		2.8
Date of Publication	2001.	2.8.6
Identifier for the Manifestation	http://ocn.loc.gov/2002598948	2.15
Identifier for the Manifestation	LC0N 2002598948	2.15
Note		2.20
Note on Title	Copyright registration information: (c) M. Michael Fauss, 1959; Bonnie Pike, 1939-	2.20.2
Note on Title	Compact disc	2.20.2
Note on Title	Brief record	2.20.2
Media Type	audio	3.2
Carrier Type	other	3.3
Number of Units 1	1 sound disc	09.rule
Dimensions	4 3/4 in.	3.5
Complete Key 1	The adventures of Pinocchio, 2001. other	

Figura 38 – Maschera d'immissione dati di un'entità di tipo *Manifestation*

Gli elementi riportati per ciascuna entità sono formulati come attributo della risorsa e ripresi direttamente dal set di elementi RDA, disponibile anche come element set RDF sull'Open Metadata Registry (<http://rdvocab.info/>). Ciascun campo è associato alla relativa istruzione di RDA Toolkit, al quale si accede previa autenticazione: si semplifica, così, l'utilizzo di uno strumento di lavoro fondamentale per le nuove generazioni di catalogatori. Le funzioni disponibili in RIMMF sono:

- creazione di nuovi modelli di catalogazione;
- creazione di nuove registrazioni catalografiche;



Figura 39 – Funzione di ricerca di dati su database esterni

- import di dati da database esterni;
- creazione di relazioni tra entità appartenenti al proprio database;
- visualizzazione della struttura gerarchica (ad albero) creata dalle relazioni tra le varie risorse.

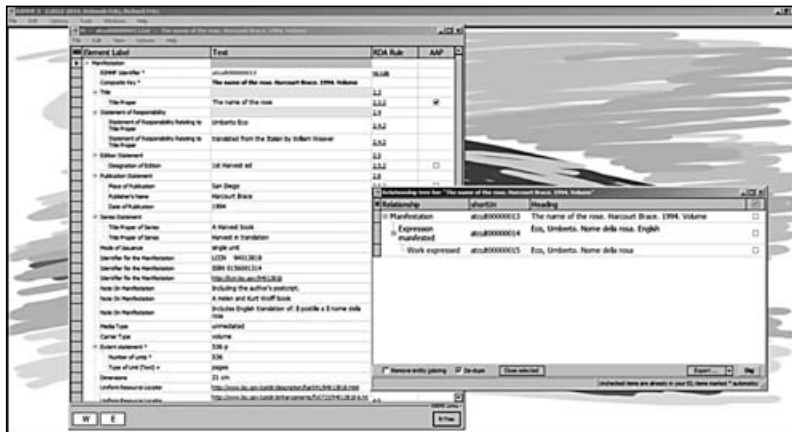


Figura 40 – Importazione di record in MARCXML dalla Library of Congress, con successiva mappatura e visualizzazione secondo la struttura RDA

Particolarmente interessante è la possibilità di visualizzare o importare in RIMMF dati provenienti da banche dati esterne: database bibliografici e OPAC (per esempio, quello della Library of Congress), authority file (per esempio, VIAF), enciclopedie e repertori online (per esempio, *Wikipedia*). Il meccanismo ha un duplice scopo:

- facilitare le operazioni di condivisione dei dati, per evitare di creare ex novo quanto già esistente su altri database;
- consentire, già in fase di strutturazione del dato, di anticipare quella che, secondo il paradigma dei linked data, è una delle attività più importanti nella strutturazione dei dati: la creazione di collegamenti esterni.

La visualizzazione in forma gerarchica delle entità descritte offre un quadro chiaro e schematizzato della complessa rete di relazioni che può essere costruita tra risorse differenti.

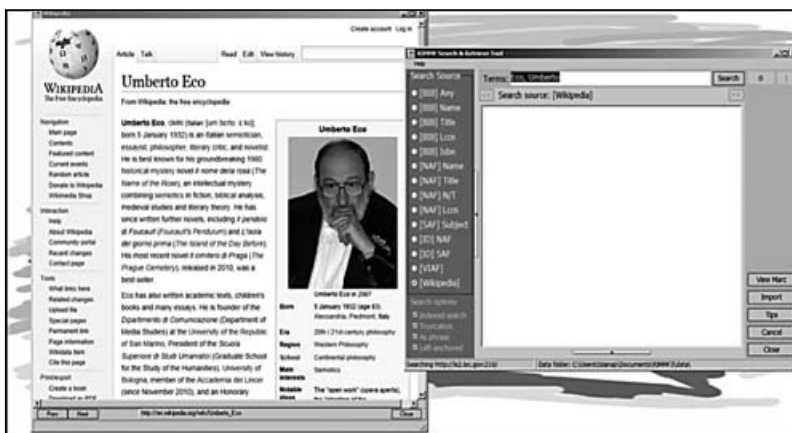


Figura 41 – Ricerca dell'entità *Umberto Eco* in Wikipedia

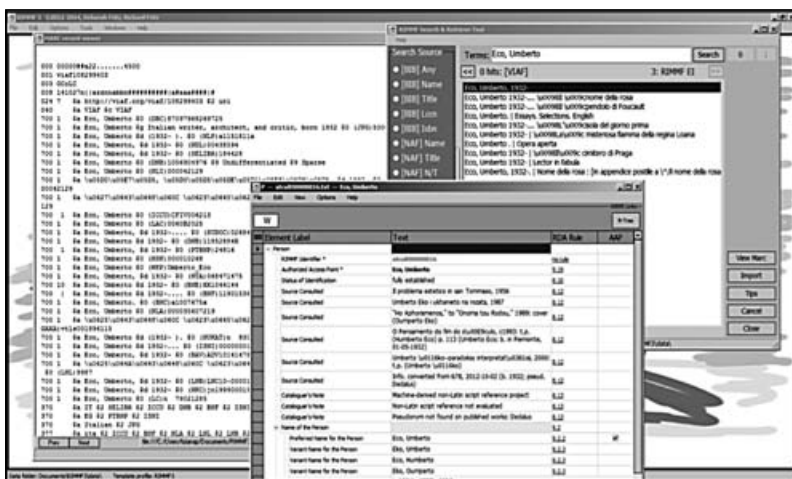


Figura 42 – Ricerca dell'entità *Umberto Eco* in VIAF, con visualizzazione del record in formato MARC21 e importazione in RIMMF

La versione consolidata di RIMMF dovrebbe garantire la possibilità di redigere descrizioni direttamente in RDA, senza passare per il formato MARC21, con la possibilità di esportare i dati in RDF, in modo da offrire un contributo diretto alla vasta comunità dei linked data.

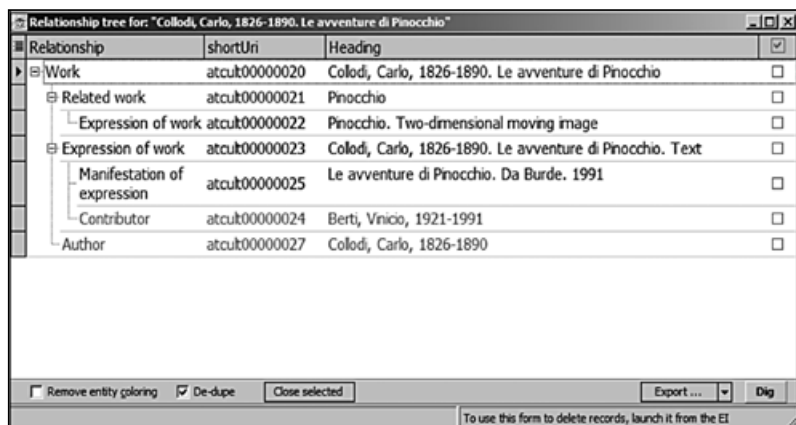


Figura 43 – Visualizzazione gerarchica delle relazioni tra entità differenti in RIMMF

8.6.2 WeCat

I software di nuova generazione devono necessariamente tener conto delle linee guida RDA, producendo strumenti che aiutino la fase di passaggio dall'epoca dominata dalla strutturazione dei dati in formato MARC, durata oltre cinquant'anni, a una nuova era, ancora non perfettamente delineata e consolidata ma che, con BIBFRAME, andrà nella direzione dei linked open data. Il primo sforzo che i nuovi software stanno realizzando è proprio quello di adeguare il linguaggio, con l'introduzione della terminologia, proposta da ICP e RDA: *punto d'accesso autorizzato* in sostituzione di *titolo uniforme; identificativo dell'opera, identificativo dell'espressione*, etc., testimonianza di un cambiamento profondo nella strutturazione del modello di metadatazione.¹² Questi cambiamenti di linguaggio nelle interfacce dei software di nuova generazione non sono ancora pienamente supportati dagli aggiornamenti del MARC21, che pure sta sostenendo uno sforzo importantissimo di evoluzione: esso, per esempio, conserva la terminologia *Titolo uniforme* per i tag 130 e 240.

¹² <www.ifla.org/files/assets/cataloguing/icp/icp_2009-it.pdf>; cfr. Mauro Guerrini, con Giuliano Genetasio, *I Principi internazionali di catalogazione (ICP)*, cit.

Nel panorama italiano, tra chi si è fatto carico di seguire e facilitare il passaggio da una dimensione legata al record alla nuova dimensione legata al dato, è @Cult che, come produttrice e distributrice di un ILS storico nel mondo delle biblioteche e degli archivi (Amicus, poi evoluto in OliSuite) ha assimilato il cambiamento che MARC21 aveva suggerito e suggerisce di continuo con i suoi numerosi aggiornamenti. Il modulo WeCat, parte dell'ILS OliSuite, è un software che, pur utilizzando ancora la registrazione in MARC21, consente di strutturare i dati secondo le linee guida RDA, in una modalità, dunque, orientata ai linked open data. Diverse sono le funzionalità che esprimono questo orientamento. Il software propone (già nella definizione degli indici di ricerca per la costruzione di query) codici specifici per le entità FRBR: la metadatozione parte con la selezione di modelli che propongono al catalogatore gli attributi previsti da FRBR per ciascuna entità specifica.

Il potenziamento delle funzioni per il controllo d'autorità esprime la necessità, comune a RDA e ai principi dei linked open data, di descrivere l'entità che si sta analizzando con un alto numero di attributi, creando così un *oggetto* perfettamente identificabile e riutilizzabile in molteplici ambiti. La possibilità di collegare le registrazioni d'autorità (per attributi di un'opera o di un'espressione) con registrazioni bibliografiche (per attributi di una manifestazione) è un elemento di grande novità, introdotto per consentire la creazione di un reticolo di collegamenti che costituisce il cuore sia di RDA sia dei linked open data. Le relazioni tra entità previste da WeCat rispecchiano le differenti tipologie

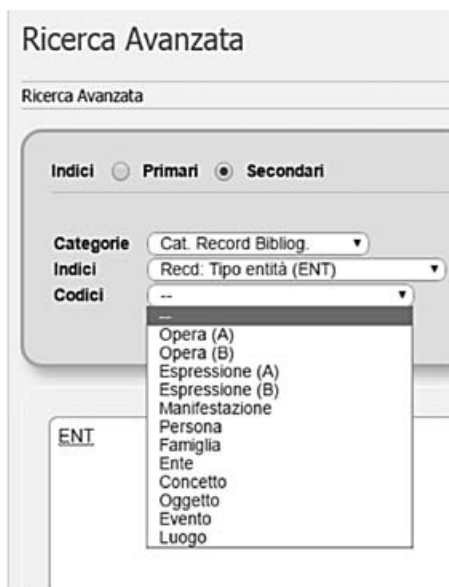


Figura 44 – WeCat: lista degli indici di ricerca per entità FRBR

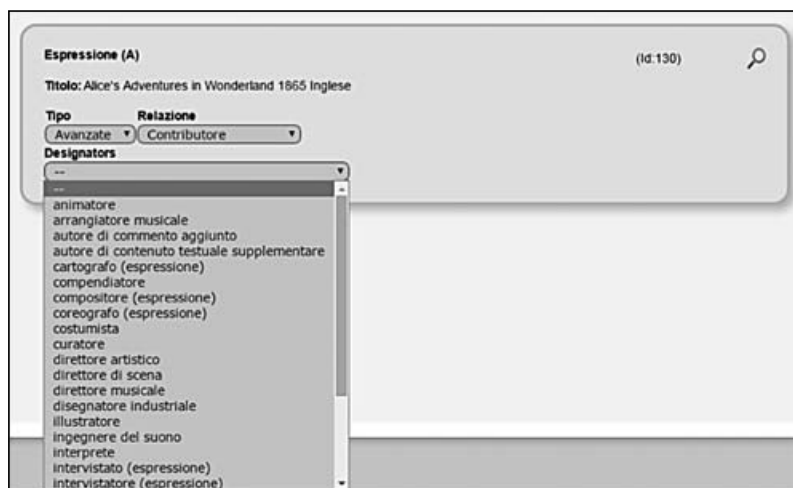


Figura 45 – Creazione di relazioni tra un'espressione e una persona (nel ruolo di contributore)

di relazione previste da RDA: le appendici I (*Relazioni tra una risorsa e persone, famiglie ed enti associate alla risorsa*), J (*Relazioni tra Opere, Espressioni, Manifestazioni e Item*) e K (*Relazioni tra persone, famiglie ed enti*) di RDA Toolkit, sono supportate nella funzione di creazione di relazione tra risorse.

Tutte le liste di termini utilizzate e proposte da WeCat provengono da vocabolari controllati, per lo più pubblicati in RDF sull'Open Metadata Registry: il software di metadatozione condivide la finalità di utilizzare un linguaggio formulato in set di

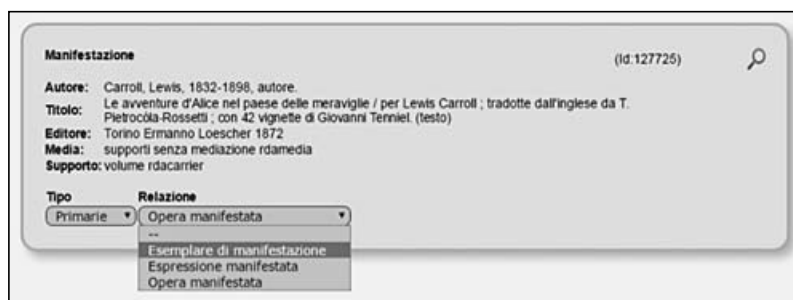


Figura 46 – Creazione di relazioni tra una manifestazione e una differente entità del gruppo 1 di FRBR

dati e in ontologie comprensibili alle macchine. Ciascuna entità, precisamente identificata, diventa, dunque, nodo di un reticolo, esteso in modo variabile, che viene mostrato graficamente per consentire al catalogatore di mantenere il controllo della rete che si sviluppa a partire da un punto di osservazione. Si veda l'esempio di relazioni tra un'opera (*Alice's Adventures in Wonderland*), le sue espressioni (le versioni del 1872 in italiano e del 1865 in inglese), il suo creatore (Lewis Carroll), le opere derivate (il film di Tim Burton del 2010).

The screenshot displays the 'Relazioni FRBR' (FRBR Relationships) interface. At the top, there's a search bar with 'Authority Records' and a breadcrumb trail: 'Ricerca Avanzata >> Record >> Relazioni FRBR'. The main section is titled 'Opera (A)' with '(Id. 129)' on the right. Below this, the 'Autore: Carroll, Lewis' and 'Titolo: Alice's Adventures in Wonderland' are listed. The 'Relazioni:' section is divided into three categories: 'Primarie' (Primary), 'Di base' (Basic), and 'Avanzate' (Advanced). Under 'Primarie', there are two entries: 'Espressione di opera:' with 'Alice's Adventures in Wonderland 1872 Italiano (Id. 131)' and 'Alice's Adventures in Wonderland 1865 Inglese (Id. 130)'. Under 'Di base', there is one entry: 'Creatore:' with 'Carroll, Lewis, 1832-1898 (autore)'. Under 'Avanzate', there is one entry: 'Relazioni derivate - basata su:' with 'Burton, Tim Alice in Wonderland (film : 2010) (Id. 133)'. Each entry has a trash icon to its right.

Figura 47 – L'opera *Alice's Adventures in Wonderland* con le sue relazioni

Nell'ottica di identificare in modo univoco le risorse, comune a RDA e ai linked open data, è interessante richiamare la funzione di ricerca e cattura, a partire da un'entità, degli identificativi utilizzati per la medesima entità su altri dataset e siti esterni, per predisporre il meccanismo d'interlinking, che consente ai dati di acquisire la quinta stella prevista dal sistema di valutazione di Tim Berners-Lee.

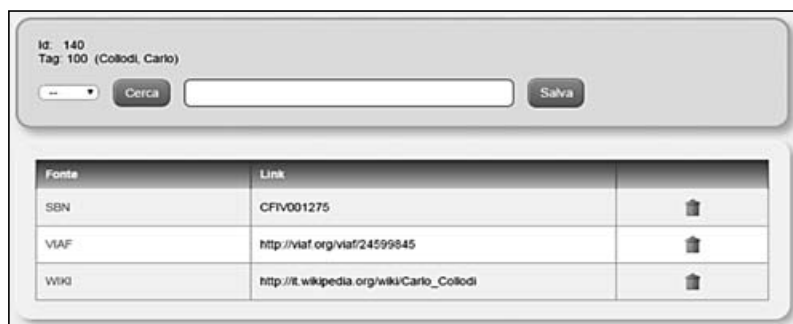


Figura 48 – Ricerca e cattura degli identificativi dell'entità persona *Collodi, Carlo* su tre fonti di dati differenti: SBN, VIAF, *Wikipedia*

Ulteriore elemento di novità di WeCat è la funzione che collega la registrazione a un framework di conversione dei dati in RDF, con l'utilizzo di differenti ontologie, frutto di un importante progetto europeo.¹³ Il collegamento tra un software di catalogazione e un convertitore di dati in linked open data ha una finalità importante: semplificare le funzioni di aggiornamento di un dataset in RDF già pubblicato nel web, consentendo a chi crea i dati di garantire un aggiornamento in tempo reale dei dati pubblicati.

¹³ Il progetto, denominato ALIADA (Automatic publication under LInked dAta paradigm of library Data), è stato realizzato da un gruppo di enti italiani, spagnoli e ungheresi: @Cult s.r.l. (I), ScanBit (S), Tecnalìa (S), ARTIUM - Centro Museo Vasco de Arte Contemporáneo (S), Szépművészeti Múzeum (U) ed è consultabile al sito <<http://www.aliada-project.eu/>>. ALIADA intende creare un framework di produzione e pubblicazione dei dati con il paradigma dei linked open data. La versione 1.0 è disponibile su GitHub, per la comunità dell'open source.

9. PROGETTI DI PUBBLICAZIONE DI DATI DELLE GRANDI BIBLIOTECHE

Le software house, che hanno il compito di accompagnare e facilitare i cambiamenti dei linguaggi e delle forme di comunicazione, stanno realizzando una nuova generazione di strumenti di elaborazione dei dati, che recepisce quanto definito dai nuovi orientamenti catalografici. L'incontro con il mondo dei linked open data, si realizza con l'adesione a RDA e con l'attenzione alle proposte di BIBFRAME. È, tuttavia, ancora presto per pensare che i dati possano essere strutturati nativamente in RDF, cioè secondo il modello definito per i linked data; ed è soprattutto alla diffusione dei dati che grandi biblioteche e importanti agenzie catalografiche del mondo hanno dedicato energie e risorse per concretizzare il modello dei linked data. Vediamo alcuni progetti significativi di pubblicazione di dati bibliografici in linked open data.

9.1 **Dati.camera.it**

Un progetto italiano molto interessante è *dati.camera.it*, che rende usufruibile il vasto e cospicuo patrimonio informativo della Camera dei deputati.¹ La pagina d'ingresso del progetto esprime la finalità del lavoro: "Conoscere la storia, l'attività, i meccanismi di funzionamento della Camera dei deputati. Condividere i dati e i documenti prodotti, disponibili al libero riutilizzo attraverso la rete". Dati.camera.it è dichiarata "una piattaforma di pubblicazione e condivisione di Linked Open Data sull'attività e gli organi della Camera, da scaricare o interrogare liberamente". Tra gli aspetti di maggior interesse, la progettazione di un'ontologia descrittiva in formato XML/RDF e uno specifico namespace OCD (Ontologia della Camera dei Deputati). Le proprietà e le classi dell'ontologia, scritta in OWL, descrivono il *dominio* Camera nel corso della sua storia e della sua attività.

¹ <<http://dati.camera.it/it/>>.

Per la conversione dei dati, il progetto utilizza e integra altre ontologie conosciute e diffuse a livello internazionale, come FOAF e Dublin Core. I dataset disponibili in linked open data sono relativi alla legislatura corrente e alle legislature precedenti, fino al 1842. I dataset storici disponibili sono organizzati per tipologia e fanno riferimento a tutte le legislature. I dataset sono scaricabili in formati differenti e sono disponibili per interrogazioni in SPARQL tramite un endpoint; sono, inoltre, fruibili utilizzando navigatori RDF disponibili sul web. Il sito riporta esempi di navigazione di entità tramite il navigatore *LodLive*.



Figura 50 – Esempio di navigazione dell'entità *Giovanni Giolitti* tramite il navigatore *LodLive*

9.2 Gruppo di lavoro Linked open data SBN

Nel gennaio 2015 l'ICCU ha pubblicato la scheda di attività del Gruppo di lavoro Linked open data SBN. Il Gruppo, costituito da rappresentanti dell'ICCU ed esperti del laboratorio VAST LAB - PIN (Polo Universitario della città di Prato - Servizi didattici e scientifici per l'Università di Firenze), ha realizzato uno schema logico e operativo di produzione e pubblicazione di un set di dati SBN strutturati in linked open data. Le attività del Gruppo si sono orientate sul lavoro di mappatura concettuale con le classi e le proprietà del modello FRBRoo (*Functional Requirements for Bibliographic Records-object oriented*), utilizzando il modello CIDOC Conceptual Reference Model (CRM), di un set di record estratto dall'OPAC SBN in formato UNIMARC.

Sulla base di tale mappatura è stata sviluppata la conversione dei dati in formato RDF ed è stato creato un prototipo di interfaccia online, su una piattaforma Aduna Sesame, per la gestione e la ricerca del set dei dati SBN in formato LOD.²

9.3 Data.bnf.fr

Il progetto data.bnf.fr della Bibliothèque nationale de France (BnF)³ utilizza i linked open data per pubblicare e rendere meglio utilizzabili i propri dati sul web; i cataloghi della BnF e i dati di Gallica e di altre biblioteche digitali.

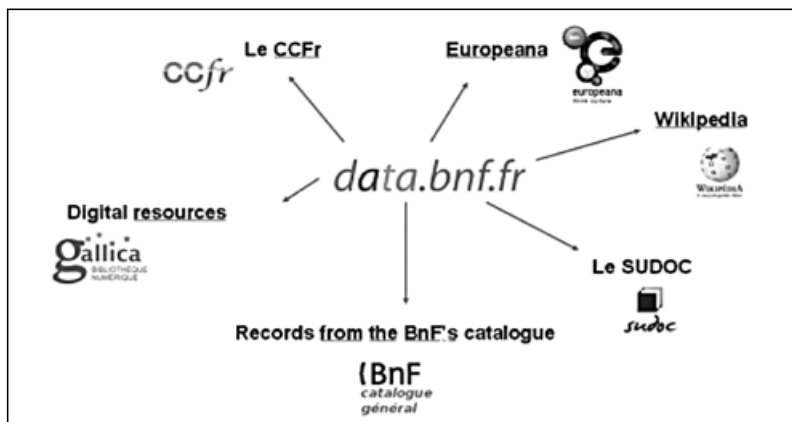


Figura 51 – Data.bnf.fr e suoi legami

Data.bnf.fr⁴ include le fasi di trasformazione dei dati originari in RDF, il loro arricchimento, il collegamento a risorse interne ed

² Il documento è accessibile sul sito dell'ICCU nella pagina web del Gruppo di lavoro LOD SBN, <http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/gruppilav_commissioni/pagina_0005.html>; il documento è consultabile all'indirizzo: <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/LOD_SBN_scheda-1.pdf>.

³ Vedi: Romain Wenz, *Linked open data for new library services: the example of data.bnf.fr*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/5509/>>.

⁴ Un'esaustiva presentazione di data.bnf.fr è disponibile sul sito del progetto: <<http://data.bnf.fr/docs/databnf-presentation-en.pdf>>.

esterne tramite interlinking, la pubblicazione in pagine HTML per la navigazione da parte degli utenti e dei motori di ricerca. Il sottotitolo del progetto è *Des fiches de référence sur les auteurs, les œuvres et les thèmes*: subito evidente è la mancanza del termine *catalogo* e l'adesione del progetto al modello FRBR, con l'attenzione del ricercatore indirizzata sulle entità principali del Gruppo 1 di FRBR (opere), del Gruppo 2 (persone e, dunque, autori) e sul Gruppo 3 (soggetti).

Per essere conforme al paradigma del linked open data, il progetto fa propri alcuni assunti fondamentali relativi alle pagine che descrivono le risorse:

- mantenere URI permanenti, strutturati in modo che siano comprensibili anche agli utenti;
- realizzare un efficiente meccanismo di negoziazione dei contenuti;
- utilizzare un modello di dati RDF compatibile, con vocabolari standard (SKOS, RDA e FOAF);

The screenshot displays the BnF data.bnf.fr interface for the entity 'Alcibiade'. The header includes the BnF logo, a search bar, and navigation links. The main content area features a book cover image, a table of metadata (Langue, Note, Domaines, Titre original), and a 'Plus de détails' button. Below this, there are tabs for 'Détails du contenu', 'Vie et éditions de l'œuvre', 'Autres pages dans data.bnf.fr', and 'Autres ressources'. The 'Vie et éditions de l'œuvre' tab is active, showing a list of documents (52) and a list of editions (5). The 'Livres' section lists the content within the work, including 'Alcibiades primus' and 'Alcibiades secundus'. The right sidebar contains a 'Outils' section with links to print, export, and report, and a 'Ressources BnF' section with links to the BnF catalog and other resources. The bottom of the page shows a 'Sites extérieurs' section with links to the European Catalogue and the OCLC WorldCat.

Figura 52 – La pagina HTML per l'entità *Alcibiade*

- utilizzare vocabolari esistenti, finché possibile;
- utilizzare vocabolario locali, creati ad hoc, solo per le classi e gli oggetti che sono specifici della biblioteca e che non troverebbero corrispondenza con i vocabolari esistenti;
- allineare i dati con dati provenienti da database esterni, dalla Library of Congress, Deutsche Nationalbibliothek, Geonames ed altri.

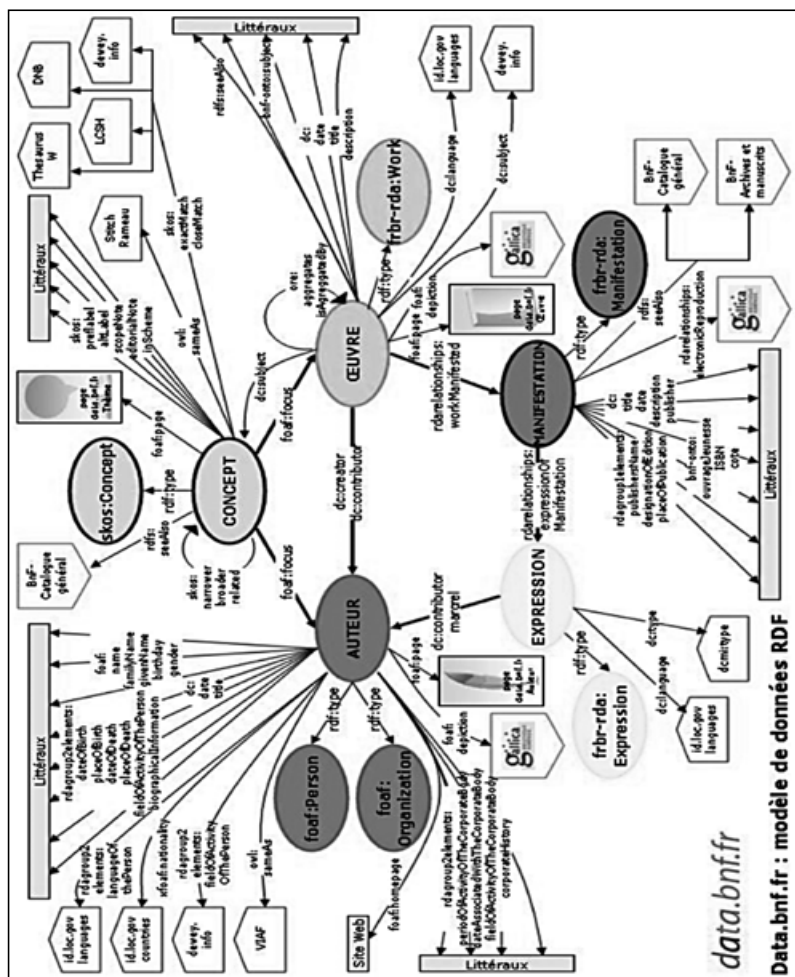


Figura 53 – Il data model del progetto data.bnf.fr ed efficace panoramica della sua architettura

Il dataset RDF, periodicamente aggiornato, è accessibile e scaricabile dalla homepage del sistema. Il data model del progetto data.bnf.fr esprime le scelte compiute: la piena adesione a FRBR, la selezione delle ontologie utilizzate, tra cui RDA (soprattutto per gli elementi di raggruppamento FRBR e, quindi, per le relazioni), FOAF (per gli elementi descrittivi le entità del tipo autore), DC (per gli elementi descrittivi della risorsa: titolo, editore, lingua, etc.), la grande attenzione rivolta alle relazioni, le cui tipologie sono tratte da RDA, i collegamenti a progetti e dataset esterni.

9.4 Bnb.data.bl.uk

Il progetto bnb.data.bl.uk della British Library pubblica una parte della Bibliografia nazionale inglese in linked open data; include pubblicazioni monografiche e seriali per un totale di circa 2,8 milioni di record. Il dataset in RDF, disponibile sotto licenza Creative Commons CC0 1.0 Universal Public Domain Dedication, è consultabile in modalità differenti: tramite un editor SPARQL e uno SPARQL endpoint per le query remote. In aggiunta, è possibile lanciare ricerche utilizzando direttamente l'interfaccia sviluppata per il progetto. Il dataset RDF è stato realizzato utilizzando una serie di schemi di dati e ontologie standard, creando nuovi termini e dichiarandoli nel British Library Terms solo nei casi di proprietà o classi non già presenti in questi schemi standard. Una specificità del progetto, per conferire una natura quanto più possibile trasversale: si è cercato, ove possibile, di scegliere schemi di dati quanto più possibile appartenenti a domini non-biblioteconomici (quali DCMI Metadata Terms, FOAF, SKOS) seppur siano presenti schemi quali ISBD elements, Bibo, RDA Group 2 Elements, Bio. I dati bibliografici sono stati analizzati e convertiti ponendo al centro del modello la risorsa, come visibile negli schemi di data model progettati per libri e seriali; da ciascuna risorsa del dataset, identificata da un URI, partono collegamenti con entità (istanze e classi) raggruppate in 4 aree:

- Autori
- Eventi di pubblicazione (nel data model dei libri, sostituito da Relazioni bibliografiche nei seriali)
- Serie
- Soggetti

Dalla stessa risorsa partono collegamenti a elementi del tipo literal, raggruppati in:

- Identificativi
- Titoli
- Informazioni miscellanee

Il dataset è stato infine arricchito di link esterni a diversi dataset, tra cui:

- GeoNames
- MARC Country code
- VIAF
- Lexvo
- LCSH
- Id.loc.gov
- Dewey

9.5 Datos.bne.es

Il progetto di pubblicazione di dati in linked open data della Biblioteca Nacional de España, ancora in versione beta, riguarda parte delle collezioni e delle risorse: libri moderni e antichi, manoscritti, mappe, disegni, stampe e fotografie, spartiti, registrazioni audio e videocassette. Il modello del progetto rispecchia il francese data.bnf.fr, a partire dalle tre tipologie di risorse trattate: Autori, Opere, Soggetti. Ciascuna risorsa ha poi un certo numero di collegamenti esterni, tra cui:

- VIAF
- Sudoc
- Dbpedia
- *Wikipedia*
- Libris
- ISNI
- Library of Congress Name Authority File
- Deutsche Nationalbibliothek
- Bibliothèque nationale de France

Il modello di dati del progetto è disegnato su FRBR; evidente anche per l'uso dell'ontologia FRBRer, formulata e approvata dall'IFLA.

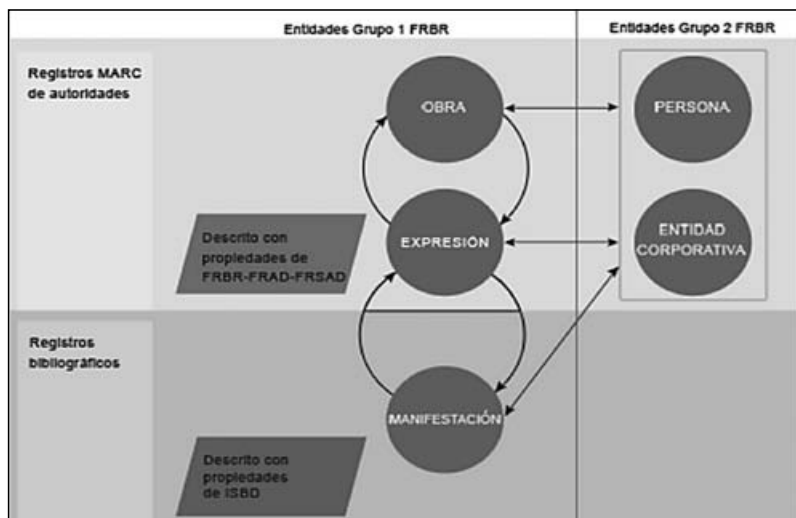


Figura 55 – Schema concettuale dell'applicazione del modello FRBR nel progetto datos.bne.es

Lo schema dei dati ISBD (ISBD elements) è molto utilizzato, insieme ad altre ontologie, tra cui: RDA, Dublin Core, e Bibo.

I dati originali del catalogo, bibliografici e d'autorità, analizzati in MARC21, sono stati utilizzati per realizzare tre mappature, necessarie per la conversione in RDF:

- la mappatura dei dati d'autorità, basata sulla combinazione di campi MARC e sottocampi per individuare le entità FRBR (Persona, Ente, Opera, Espressione);
- la mappatura delle relazioni esistenti tra le entità, analizzate sia tra record d'autorità sia tra record bibliografici;
- la mappatura di ciascun campo e sottocampo MARC con le proprietà delle ontologie selezionate.

9.6 Il linked data service della Deutsche Nationalbibliothek

La Deutsche Nationalbibliothek (DNB) ha avviato nel 2010 il progetto di conversione e pubblicazione dei propri record d'autorità in linked open data;⁵ nel 2012 ha esteso il progetto ai

⁵ <<http://www.dnb.de/EN/lds>>.

record bibliografici. A gennaio 2015 ha pubblicato in RDF, sotto licenza CC0, i dati delle principali collezioni (a eccezione della musica a stampa e della Exil Sammlungen) e dei seriali della Zeitschriftendatenbank (ZDB). I record bibliografici e d'autorità sono stati analizzati nei loro formati originari per individuare un sottoinsieme di elementi necessari all'identificazione delle risorse. L'analisi ha seguito le raccomandazione⁶ del DINI WG KIM Bibliographic Data Group,⁷ costituito dai rappresentanti delle reti e delle grandi biblioteche di area linguistica tedesca. Il gruppo di lavoro ha definito criteri di armonizzazione delle rappresentazioni bibliografiche in RDF valide per tutti i paesi di lingua tedesca. Nella modellazione dei dati sono stati utilizzati vocabolari e ontologie esistenti e ampiamente utilizzati in progetti analoghi, tra cui:

- Dublin Core terms
- Dublin Core elements 1.1
- Bibliographic Ontology (Bibo)
- Resource Description and Access Vocabularies
- ISBD
- RDF Schema
- Upper Mapping and Binding Exchange Layer (Umbel)⁸
- FOAF
- Schema.org: Library extension terms⁹

È evidente l'influenza di RDA, per esempio, nell'attenzione rivolta agli elementi quali *content*, *media* e *carrier*, nella definizione degli attributi qualificanti e identificanti la risorsa.

I dati pubblicati in linked open data sono accessibili in diverse modalità:

- tramite lo specifico URI della risorsa;
- come dataset RDF;
- tramite interfacce SRU e OAI-PMH.

⁶ <<https://wiki.dnb.de/pages/viewpage.action?pageId=68060017>>.

⁷ <<https://wiki.dnb.de/display/DINIAGKIM/Titeldaten+Gruppe>>.

⁸ <<http://umbel.org/umbel#>>.

⁹ <<http://purl.org/library/>>.

KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK

Gesamter Bestand Musikharchiv Exilsammlungen Buchmuseum

→ Suchformular zurücksetzen

 Finden → ☒ Expertensuche ?

Ergebnis der Suche nach: "bertolt" and "brecht"

[← Zurück zur Trefferliste](#)

Treffer 2 von 6278

Link zu diesem Datensatz	http://d-nb.info/1018884890
Titel/Bezeichnung	Bertolt Brecht, Der gute Mensch von Sezuan, Parabelstück ... verstehen / erarb. von Stefan Volk, Hrsg. von Johannes Diekhans ; Michael Völd
Person(en)	Volk, Stefan
Ausgabe	1. Dr.
Verleger	Paderborn : Schöningh
Erscheinungsjahr	2015
Umfang/Format	205 S. : Ill. ; 19 cm
ISBN/Einband/Preis	978-3-14-022523-6 kart. 3-14-022523-7
LAN	9783140225236

Aktionen

- In meine Auswahl übernehmen
- Druckansicht
- MARC21-XML-Repräsentation dieses Datensatzes
- RDF/XML-Repräsentation dieses Datensatzes
- Dokumentation Linked Data
- bE: BIFRAME-Repräsentation dieses Datensatzes
- Projekt BIFRAME
- Korrekturanfrage

☒ Lieferbar bei buchhandel.de?

[Teilen](#)

[Tweet](#)

Figura 56 – Record del catalogo della Deutschen Nationalbibliothek con la funzione, nella sezione destra, di scarico in diversi formati, tra cui RDF/XML

9.7 RDA come istruzione e linked open data come realizzazione

Dalla breve analisi dei più importanti progetti di pubblicazione di dati in linked open data, in campo bibliografico, diventati modello di riferimento per altri progetti, è possibile trarre qualche riflessione:

- il modello FRBR è fortemente riconosciuto in tutti i progetti, seppure non sempre realizzato. Nella documentazione dei progetti vi è sempre il riferimento alle entità di FRBR (per esempio, data.bnf.fr), o la dichiarazione di utilizzo di un'ontologia nata dal modello;
- scopo primario dei progetti è la realizzazione di una struttura che sfrutti al massimo i dati esistenti e li colleghi ad altri dataset per l'arricchimento informativo, realizzato tramite l'interlinking;
- la filosofia che sottende il trattamento del dato nei progetti bibliografici è la medesima (il paradigma dei linked data e del web semantico) e in molti casi anche i dati collegati coincidono in parte. Ciò che cambia e caratterizza un progetto rispetto a un altro è la modalità di rappresentazione dei dati:

- un'attenzione particolare è rivolta ai fruitori dei progetti. Le interfacce di fruizione e di visualizzazione dei dati cambiano in funzione di una presunta loro maggiore facilità di lettura;
- nella realizzazione dei progetti pesano moltissimo gli investimenti di risorse economiche e umane che ciascun ente può realizzare.

La filosofia da adottare (il modello), la tecnologia e la consapevolezza della strada da intraprendere sono chiare. La qualità dei risultati dipende dalla quantità di investimenti economici e umani. Ciò dimostra quanto afferma RDA:

- identificare e collegare i dati attiene alla teoria della catalogazione, acquisita e consolidata;
- la traduzione delle realizzazioni in progetti di fruizione, in primis la rappresentazione, attiene al grado e alla quantità di investimenti di risorse finanziarie e umane.

Potremmo concludere che RDA è la faccia normativa di ciò che i linked open data realizzano concretamente.

9.8 ALIADA: un framework di pubblicazione di linked open data

Il progetto¹⁰ ALIADA (Automatic publication under LInked dAta paradigm of library Data),¹¹ approvato e finanziato dall'Unione europea all'interno del Seventh Framework Programme for Research (FP7), prevede la collaborazione tra cinque partner – @Cult (I), ScanBit (S), Tecnalia (S), ARTIUM - Centro Museo Vasco de Arte Contemporáneo (S), Szépművészeti Múzeum (U) – dislocati in tre nazioni europee: Spagna, Italia e Ungheria. I primi tre rivestono un ruolo tecnico, con la responsabilità di gestire e creare gli artefatti, per esempio, software e documentazione. Gli altri due partner sono istituzioni culturali che rappresentano i bisogni dei fruitori, fornendo suggerimenti e requisiti funzionali da parte di chi conosce il “dominio culturale”. Obiettivo del progetto è la realizzazione di un framework di conversione e

¹⁰ Il paragrafo è redatto da Andrea Gazzarini.

¹¹ <<http://aliada-project.eu>>.

pubblicazione automatica in RDF dei dati appartenenti a istituzioni culturali. La procedura di conversione non è strettamente legata a un dominio; tuttavia, il progetto definisce una serie di formati in ingresso – per esempio, LIDO (Lightweight Information Describing Objects), MARCXML, Dublin Core – e la corrispondente caratterizzazione dei dati in uscita. Il piano di progetto prevede diverse iterazioni, per un ciclo di sviluppo complessivo di circa due anni che terminerà nella seconda metà del 2015. Periodicamente vengono resi disponibili i prototipi legati alle *milestone* (punti di verifica) di progetto: ognuno di essi assicura una percentuale di copertura dei requisiti funzionali totali.

ALIADA viene rilasciato sotto licenza GPLv3 ed è, pertanto, un progetto open source: il codice sorgente e la documentazione relativa possono essere prelevati e utilizzati dalla comunità del web. È, quindi, possibile, anzi consigliato partecipare, in varie forme e modalità, alla disseminazione del progetto: stesura di documentazione, partecipazione ai forum, alle mailing list, allo sviluppo di funzionalità, apertura di segnalazioni per nuove richieste o malfunzionamenti. Il progetto, nell’ottica di aumentare il grado di coesione ed estendibilità, è stato strutturato in diversi moduli, ognuno dei quali assolve a uno specifico compito nella catena complessiva di lavoro che parte dalla gestione dei dati di input sino alla loro pubblicazione, convertiti in *tuple* RDF.

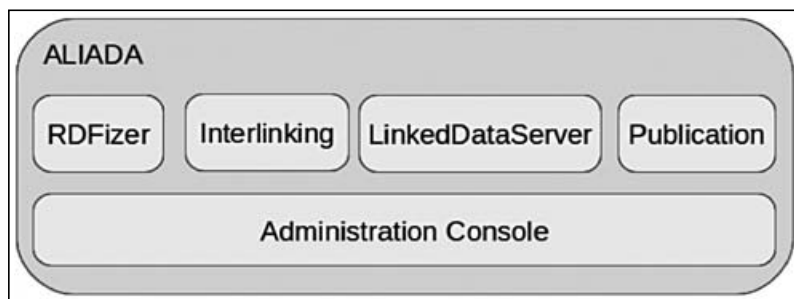


Figura 57 – Moduli che compongono ALIADA e le loro relazioni

9.8.1 La consolle di amministrazione

Il processo di conversione e pubblicazione dei dati rappresenta il fattore chiave di ALIADA; la consolle di amministra-

zione, gestita da un utente professionale, consente il controllo delle funzioni: caricamento dei dati, visualizzazione e monitoraggio dei processi. L'applicazione è stata disegnata cercando di astrarre e semplificare la complessità delle operazioni che vengono svolte dal sistema. L'interfaccia di amministrazione permette, dunque, di gestire i dati di input, produrre dei template (modelli) preconfigurati per definire le regole di conversione, lanciare i processi di trasformazione e monitorarne lo status (andamento).

9.8.2 RDFizer: modulo di conversione

Alla base della catena di produzione del dato RDF si trova il modulo di conversione, denominato RDFizer. Il suo ruolo è molto semplice: prendere un insieme di dati in input, in un formato supportato, applicare regole di trasformazione e infine produrre un output. Il modulo è progettato per trattare dati in qualunque formato di input e di output. La natura modulare lo rende conforme al principio Open-Closed,¹² secondo il quale un sistema deve poter permettere l'aggiunta di nuove funzionalità, fornendo punti di estensione ed evitando effetti indesiderati relativi all'interruzione di funzionalità esistenti. Il modulo opera la conversione dei dati di input in tuple RDF, in accordo con le ontologie definite per il progetto.

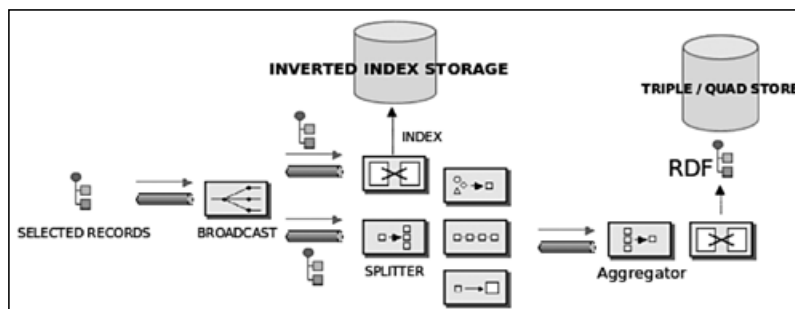


Figura 58 – Schema del funzionamento, ad alto livello, del modulo di conversione

¹² Bertrand Meyer, *Object-Oriented software construction*. Second edition. Prentice Hall PTR. Upper Saddle River, New Jersey, 1997, p. 57.

A sinistra dello schema vi sono i dati di input, a destra la produzione di dati RDF, con conseguente memorizzazione all'interno dello store, il database che accoglie nativamente strutture RDF. Nel mezzo, i componenti formano una catena asincrona che suddivide il processo di conversione in maniera parallela, di modo da aumentare la produttività complessiva del sistema in termini di tempi di risposta. Il cuore funzionale di RDFizer è costituito dai cosiddetti *template di conversione*, ovvero file in cui l'utente definisce le regole di conversione tramite uno pseudo-linguaggio ad alto livello. Di seguito un estratto di un template di conversione:

```
$work $is_a #efrbroo("F1_Work") .  
$title $is_a #efrbroo("E35_Title") .  
$title $P3_has_note "La Divina Commedia" .  
$work $P102_has_title $title .
```

Le espressioni che definiscono le regole di conversione sono intuitive, molto semplici da utilizzare, creare e mantenere per una persona formata sui meta-linguaggi di programmazione. All'interno della catena di conversione, ogni volta che viene identificato un record in input, il sistema associa il record a un template di conversione; le regole del template vengono applicate e le tuple RDF sono così prodotte in uscita.

9.8.3 Interlinking

La produzione di tuple RDF a partire da un formato in ingresso soddisfa un requisito di standardizzazione e aumenta la caratteristica di riusabilità del dato. La reale potenza di questo formato di dati è, tuttavia, rappresentato dalla possibilità di creare legami con concetti ed entità presenti su dataset esterni. A fronte del grafo RDF prodotto da RDFizer, il modulo dell'interlinking, tramite un set di regole configurabili da interfaccia di amministrazione, provvede a cercare entità in dataset esterni e a creare collegamenti. Il risultato è una serie di triple con:

- soggetto: gli URI delle entità locali ricavate dal processo di conversione;
- predicato: *owl:sameAs* o una sua sotto-proprietà;
- oggetto: l'URI della risorsa esterna collegata.

A questo punto RDF store contiene un dataset interrogabile (per esempio, SPARQL) e arricchito dai collegamenti esterni.

9.8.4 *LinkedDataServer*

Oltre alle possibilità d'interrogazione offerte da RDF store, il quale funge da SPARQL endpoint, il modulo espone un set di API REST che permettono di dereferenziare gli URI delle entità, restituendo una loro rappresentazione. La tipologia della rappresentazione, il formato, dipende dalla negoziazione con il client, sulla base di attributi presenti nella richiesta.

9.8.5 *Pubblicazione*

Convertiti in RDF, arricchiti con l'interlinking e disponibili sul LinkedDataServer, i dati potrebbero essere qualificati come *pubblici*, purché l'accesso SPARQL e LinkedDataServer lo siano. Il modulo di pubblicazione è finalizzato a estendere il concetto di disponibilità e fruizione dei dati, favorendo l'automazione del processo di consegna e la disponibilità del dataset generato su piattaforme pubbliche come datahub.io.¹³ Il processo, in parte automatico e in parte manuale, prevede l'aderenza a una serie di principi formali, quali per esempio, consistenza del dataset, URI dereferenziabili, presenza di un LinkedDataServer che supporti la negoziazione del contenuto. ALIADA aiuta l'utente automatizzando i passaggi che possono essere svolti dalla macchina e fornendo meccanismi di controllo ausiliari per le operazioni che devono essere svolte manualmente.

9.8.6 *Tipologie di dati di input*

Le tipologie di dati di input previste (al gennaio 2015) dal sistema corrispondono ai formati adottati dalle istituzioni partner del progetto:

- MARCXML:¹⁴ utilizzato come formato di rappresentazione di record bibliografici e d'autorità;

¹³ <<http://datahub.io/>>.

¹⁴ MARC21 XML Schema, <<http://www.loc.gov/standards/marcxml/>>.

- LIDO:¹⁵ utilizzato per i dati museali;
- Dublin Core:¹⁶ per la rappresentazione di materiale digitale.

Il progetto, tuttavia, essendo strutturato come framework, può essere facilmente esteso per aggiungere nuovi formati di input e di output.

9.8.7 *L'ontologia*

La prima fase del progetto ha analizzato numerose ontologie su cui modellare il modulo di conversione. Il risultato dell'analisi è una nuova ontologia, consultabile all'indirizzo <http://www.aliada-project.eu/2014/aliada-ontology.owl>.

Essa include altre ontologie individuate e scelte per rappresentare le entità trattate:

- Erlangen FRBRoo (EFRBRoo¹⁷): basata sulla definizione dell'ontologia FRBRoo,¹⁸ aggiunge e combina elementi proveniente dalle ontologie ECRM¹⁹ e CIDOC-CRM;²⁰
- WGS84 Geo Positioning:²¹ per rappresentare concetti geospaziali, localizzabili tramite coordinate geografiche;
- OWL-Time:²² per i concetti a connotazione temporale;
- SKOS-XL (Simple Knowledge Organization System eXtension for Labels):²³ estensione di SKOS,²⁴ che aggiunge entità a concetti che nella versione base sono rappresentati da literal;
- FOAF:²⁵ per rappresentare concetti relativi a persone, organizzazioni, enti e corrispondenti relazioni.

¹⁵ Lightweight Information Describing Objects Version 1.0, <<http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>>.

¹⁶ Dublin Core, <http://it.wikipedia.org/wiki/Dublin_Core>.

¹⁷ <<http://erlangen-crm.org/efrbroo>>.

¹⁸ <http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0.2.pdf>.

¹⁹ <<http://groups.google.com/group/erlangen-crm>>.

²⁰ <<http://www.cidoc-crm.org/>>.

²¹ <http://www.w3.org/2003/01/geo/wgs84_pos>.

²² <<http://www.w3.org/2006/time>>.

²³ <<http://www.w3.org/TR/skos-reference/skos-xl.html>>.

²⁴ <<http://www.w3.org/TR/skos-primer/>>.

²⁵ <<http://www.foaf-project.org/>>.

L'ontologia di ALIADA è, dunque, una meta-ontologia, che definisce proprie entità, che derivano dai sistemi di classificazione usati dai partner coinvolti nel progetto.

Esempio:

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix aliada: <http://aliada-project.eu/2014/aliada-ontology/>.

<http://aliada-project.eu/2014/aliada-ontology/resource/
Concept/VIAFIdentifier>
  a <skos:Concept>;
  <rdfs:comment> "VIAF identifier used by LIDO for
Actor descriptions.";
  <skos:prefLabel> "VIAFIdentifier"@en;
  <skosxl:literalForm> "VIAFIdentifier"@en.
```

9.8.8 *Evoluzioni*

ALIADA si pone due obiettivi: il primo riguarda i requisiti funzionali definiti in ambito europeo, per i quali è stato finanziato nell'ambito dell'FP7, e da realizzare entro il 2015; il secondo, con un orizzonte temporale a lungo termine, riguarda la diffusione e la condivisione del progetto, stimolata dalla partecipazione e dal contributo dei soggetti coinvolti: istituzioni, tecnici, interessati a vario titolo. Per queste ragioni il progetto ha previsto l'estensione di moduli per favorire funzionalità aggiuntive e rispondere a esigenze nuove o particolari come: l'uso di dati di input e output di tipologie differenti; un gestore di ontologie in un'area dedicata nella consolle di amministrazione in cui l'utente può raccogliere, censire e gestire le ontologie utilizzate dal sistema nella conversione. Altro ambito di miglioramento è l'adozione di un linguaggio più vicino all'utente funzionale per la stesura delle regole di conversione.

Il linguaggio potrebbe essere semplificato con l'utilizzo di widget grafici (icone), ovvero di un'interfaccia grafica che con-

senta all'utente di costruire espressioni tramite funzioni quali *drag & drop*,²⁶ anziché digitare le espressioni di conversione.

ALIADA è un progetto open source e in quanto tale si affida alla partecipazione e al contributo di idee da parte della comunità degli utenti.

²⁶ *Drag-and-drop* indica una successione di tre azioni, consistenti nel cliccare su un oggetto virtuale (quale una finestra o un'icona) per trascinarlo (in inglese: *drag*) in un'altra posizione, dove viene rilasciato (in inglese: *drop*) (*Wikipedia*).

10. LINKED DATA: BENEFICI PER GLI ENTI E PER GLI UTENTI

L'adesione degli istituti della memoria registrata al processo che li renda parte integrante del web semantico può essere realizzata in fasi successive, partendo dalla conversione di set di dati esistenti in RDF, fino alla realizzazione di progetti complessi, che prevedano meccanismi di interlinking e la produzione di dati in sintassi RDF nativi. Il cammino delineato offre vantaggi considerevoli agli enti che pubblicano i loro dati e agli utenti che accedono a essi.

La creazione di contenuto distribuito: le tecnologie del web semantico consentono d'aggregare dati eterogenei prodotti da enti differenti in portali globali, con costi molto contenuti rispetto alle strategie tradizionali (che prevedono per lo più l'uso di content management system e di elaborati meccanismi d'aggregazione e di aggiornamento dei dati) e con risultati estremamente nuovi e interessanti. La possibilità di collegare tramite link interni e esterni il dato sul manoscritto o su un articolo di un autore locale, con la biografia e l'identificatore dell'autore (su *Wikipedia*, VIAF, ORCID, ISNI, ResearcherID), con un saggio sul manoscritto presente sul Catalogo bibliografico, con il carteggio tra questo autore e altre personalità descritto in un'altra banca dati, e con la bibliografia prodotta da un ente esterno, arricchisce il dato iniziale con una serie di dati esterni autorevoli che altri enti mettono a disposizione nel web. La concezione di contenuti digitali distribuiti e interconnessi in rete "ha, in ultima analisi, lo scopo di formare una rete internazionale di repository indicizzati ricercabili da un'unica interfaccia".¹

La gestione automatica dei link e degli URI che identificano le risorse: i meccanismi d'identificazione delle risorse nel web

¹ University of Oregon Libraries, *Statement*, <http://library.uoregon.edu/diglib/irg/SB_Role.html>. Nel 2013 NISO ha lanciato un'iniziativa per sviluppare standard per i "metadati e indicatori open access", particolarmente per i dati bibliografici e gli identificatori relativi all'accessibilità degli articoli di rivista, nel rispetto delle condizioni della loro politica editoriale; vedi: <<http://www.niso.org/publications/newsline/2013/newslinefeb2013.html#report2>>.

e di collegamenti ad altre risorse sono onerosi se risolti con le tecnologie del web tradizionale; il rischio più prossimo è l'obsolescenza e il conseguente puntamento cieco a risorse inesistenti o diversamente identificate. La creazione e la gestione degli URI persistenti e dei collegamenti previste dal web semantico in modalità automatica, basata su metadati e ontologie condivise, garantisce la permanenza e l'aggiornamento di questi importanti identificativi e dei link che li collegano nel web.

La condivisione di strumenti di pubblicazione dei contenuti: nelle modalità tradizionali di pubblicazione dei contenuti ciascun ente deve provvedere a proprie spese alle attività di pubblicazione e di promozione dei dati; le tecnologie del web semantico indicano, invece, modalità di condivisione nella esposizione dei propri dati,² a partire, per esempio, da strumenti open source quali CKAN Data Hub.³ CKAN è un potente data management system che rende accessibili i dataset fornendo agli enti interessati strumenti per la loro pubblicazione, condivisione, ricerca e fruizione. I progetti di conversione in linked data prevedono in genere, in una fase operativa finale, la pubblicazione del dataset in RDF su Data Hub, che offre agli editori una serie di servizi, tra cui un'interfaccia di ricerca e la possibilità di scaricare dataset d'interesse e di condividere i dati con altri enti.

L'arricchimento semantico delle risorse: il meccanismo dell'interlinking arricchisce automaticamente e senza costi aggiuntivi il dataset originario; allo stesso tempo diventa strumento di arricchimento per chiunque (enti esterni) crei il medesimo meccanismo di collegamento verso questi dati, proprio per il fenomeno di generazione di processi virtuosi di arricchimento reciproco tra i partecipanti innescato dall'interlinking.

Riutilizzo dei dati prodotti: il concetto di *riutilizzo del dato* è uno dei fondamenti del web semantico, e prevede appunto che il dato creato o convertito in RDF possa essere utilizzato in contesti e situazioni differenti senza apportarvi alcuna modifica; questo potenziale di riutilizzo a costo zero del dato non è del

² Per la promozione di library linked dataset, vedi: Library Linked Data - the Datahub, *Library Linked Data Incubator Group Final Report*.

³ <<http://okfnlabs.org/blog/2014/09/11/data-api-for-data-packages-with-dpm-and-ckan.html>>.

web tradizionale, in cui è difficile immaginare di costruire nuovi servizi senza prevedere un'elaborazione del dato originario.

I benefici più evidenti per gli utenti sono:

Fruizione integrata e globale di risorse distribuite: consente di estendere virtualmente il dataset che si sta esplorando verso dataset esterni, a partire da quelli più comunemente utilizzati come *Wikipedia*, *GeoNames* e *VIAF*.

Ricerca semantica: nei portali, negli OPAC, nei siti tradizionali le possibilità di ricerca, per quanto potenziate dai motori e dai discovery tool, sono limitate a funzioni come l'interrogazione per parola-chiave, per indici, per strutture gerarchiche (tipo sistemi di classificazione, thesauri, soggettari) e poco altro. I sistemi semantici potenziano enormemente le possibilità di ricerca, sfruttando al massimo questi strumenti e consentendo l'attivazione di processi inferenziali che accrescono il potere informativo dei dati.

Costruzione di servizi aggiuntivi come quello di recommendation system: la pubblicazione dei propri dati in open linked data consente a chiunque di costruire servizi basati, per esempio, sulle consultazioni di dati compiute da altri utenti per suggerire nuovi percorsi di ricerca o percorsi alternativi a quelli tradizionali: consultazioni sfociate in blog, commenti, social metadata, funzioni Condividi/Share su *Facebook*, *Twitter*, *Mendeley*, *LinkedIn*, *Research Gate*.

La comunità del web semantico ha intuito da tempo il prezioso valore che la comunità dei beni culturali, con la sua lunga e profonda esperienza e attenzione alla costruzione di una informazione ricca, qualitativamente controllata, strutturata e condivisibile, può apportare al patrimonio informativo del vasto spazio globale costituito dal web e dalla sua popolazione di utenti. Che adesso la consapevolezza dell'importanza di questa collaborazione, senza chiusure a comunità differenti, sia reciproca è già un enorme passo avanti: la strada per realizzare l'integrazione è tracciata e in alcuni casi realizzata; di conseguenza, presentare una solida e fertile base per la creazione dell'ecosistema Linked Open Data repository, che ingloberà metodologie e sistemi di LODLAM (*Linked Open Data in Libraries, Archives, and Museums*) e *Linked Open Data research* repository distribuiti a livello globale in rete.

Stiamo vivendo una situazione epocale in cui l'introduzione di una tecnologia incide profondamente nel *modus operandi* di un'ampia comunità, determinando un nuovo modo d'intendere, di costruire e di condividere l'informazione. Le biblioteche, gli archivi, gli istituti della memoria e della cultura non possono prescindere da questa tendenza e da questo nuovo scenario tecnologico, per essere sempre più e sempre meglio parte di una più vasta, eterogenea e ricca comunità globale.⁴

⁴ Hanno, per esempio, espresso analogo orientamento gli storici; vedi: Roberto Delle Donne, *Un intreccio di iniziative scientifiche. Reti medievali e il futuro della storiografia digitale*, "Reti medievali rivista", vol. 15, n. 2 (2014), p. 93-156, <<http://rivista.retimedievali.it>>.

DAI CATALOGHI ALLA NAVIGAZIONE SEMANTICA

di Carlo Bianchini

1. La scienza degli indici

La Bibliografia, secondo Alfredo Serrai, è la “scienza degli indici”, dove per *indice* s’intende qualcosa che serve a indicare e, per estensione, un elenco, un catalogo ovvero di una lista ordinata e sistematica di più oggetti della stessa specie. L’indice-elenco offre al suo lettore un surrogato delle cose o delle persone che rappresenta; esso descrive un insieme A di oggetti definito e organizzato in base a un criterio. Infatti, una definizione corretta di bibliografia – intesa come repertorio bibliografico – è “una lista di libri o documenti ordinati secondo un’esigenza sistematica”.¹ L’indice-elenco che si ottiene dal processo di allestimento di un repertorio bibliografico è, concretamente, una sequenza di dati, che si ripetono secondo uno schema (stile di citazione) e che indicano gli attributi degli oggetti che appartengono ad A. Per esempio, in un elenco che rappresenta un insieme A di persone così definito:

$$A = \{\text{Tizio, Caio, Sempronio}\}$$

le persone sono rappresentate tramite un particolare attributo: il nome. Nell’esempio lo stile di citazione è ridotto al minimo, dato che prevede il solo nome in carattere tondo.

Il nome specifico di una persona ha una funzione identificativa: Tizio, Caio e Sempronio possono essere distinti e identificati proprio grazie al riferimento al loro nome specifico. Non tutte le caratteristiche delle persone sono altrettanto identificative: per esempio, un altro attributo importante delle persone, come il sesso, non sarebbe stato adatto a svolgere la funzione di identificazione, dal momento che tutti gli elementi dell’insieme A appartengono allo stesso sesso. Se l’elenco si allunga, perché rappresenta un insieme più grande di persone, è possibile che il

¹ Attilio Mauro Caproni, *Bibliografia generale*. Napoli: Società Editrice Napoletana, 1980, p. 7-8.

nome di battesimo diventi, da solo, insufficiente a identificare in modo certo (per esempio, il riferimento a Sempronio diventa inadeguato se le persone che portano quel nome sono due); per questo motivo, è necessario usare l'attributo di partenza (il nome) assieme a un altro attributo (comunemente, il cognome), come in figura 1.

Francesco Bianchi
Nadia Bruni
Teresa Mori
Francesco Rossi
Giacomo Verdi

Figura 1 – Elenco

Con l'aumento del numero di elementi che costituiscono l'insieme oggetto dell'indice, aumenta progressivamente il numero degli attributi richiesti per giungere all'identificazione di una persona. Di norma, quando si compila un modulo per una pubblica amministrazione, i dati richiesti per garantire l'identificazione della persona sono: nome, cognome, luogo di nascita (con l'aggiunta della Provincia o dello Stato), data di nascita, dati completi di residenza e, spesso, il codice fiscale, come in figura 2.

Nome: Francesco	Cognome: Bianchi
Sesso: M	
Data di nascita: 01/04/1980	Luogo di nascita: Faedis
Provincia: UD	
C.F.: BNCFNC80D01D455X	

Figura 2 – Modulo

Le due figure consentono di mettere in evidenza alcune caratteristiche fondamentali dei dati necessari per la creazione degli elenchi. Prima di tutto, il *numero* degli attributi necessari per identificare in un elenco un elemento di un insieme dipende dalle dimensioni dell'insieme che si indicizza. In secondo luogo, la *scelta* degli attributi utilizzati per indicare gli elementi dell'elenco dipende dalla capacità distintiva di ciascun attributo. Per esempio, nel caso della figura 2, per *descrivere* la perso-

na potrebbero essere sufficienti alcuni attributi; per *identificare* la persona invece sono necessari in alternativa tutti gli attributi delle prime due righe, oppure un solo attributo che funge da identificatore (nell'esempio è il codice fiscale che, come è noto, si calcola proprio a partire da tutti gli attributi precedenti). Questa alternativa consente di evidenziare un'altra importante caratteristica dei dati necessari per allestire un elenco-indice: un elemento di un insieme, anche grande, si può identificare tramite un'accurata descrizione (per esempio, nome, cognome, sesso, data e luogo di nascita) o tramite un dato identificativo (per esempio, il codice fiscale).

La funzione di identificazione è essenziale negli indici, ma non è l'unica. Per essere efficace, un indice dev'essere organizzato, per consentire percorsi di ricerca basati anche su caratteristiche o attributi diversi da quelli utilizzati per l'identificazione. L'indice deve avere, quindi, una struttura sindetica, ovvero di connessioni tra i dati relativi agli elementi dell'insieme che consenta di navigare attraverso l'insieme. Identificazione e organizzazione degli elementi in base a caratteristiche scelte per la loro natura funzionale sono necessità che corrispondono a due possibili approcci nell'utilizzo dell'indice: la *ricerca consapevole*, ovvero nella quale l'utente sa esattamente ciò di cui ha bisogno e parte da elementi noti, e la *ricerca esplorativa*, nella quale l'utente non conosce affatto o non è in grado di esprimere l'oggetto di interesse, che si avvale della *navigazione*.²

2. I metadati: una definizione funzionale

Una distinzione importante tra le registrazioni di figura 1 e figura 2 è che in figura 1 non è specificato che “Francesco” è un nome e “Bianchi” è un cognome. Infatti un italiano che legge

² Elaine Svenonius, *The Intellectual Foundation of Information Organization*, Cambridge, Mass.: MIT Press, 2000, p. 19; Trad. it.: Elaine Svenonius, *Il fondamento intellettuale dell'organizzazione dell'informazione*, a cura di Maria Letizia Fabbrini, Firenze: Le Lettere, 2008, p. 28; Mauro Guerrini, Carlo Bianchini, Andrea Capaccioni, *La biblioteca spiegata agli studenti universitari*, Milano: Editrice Bibliografica, 2012, p. 82.

l'elenco è perfettamente in grado di riconoscere i due dati per ciò che sono. In figura 2 invece, il tipo di dato è stato specificato per consentire al lettore-compilatore di inserire i dati corretti nel luogo appropriato del modulo. In entrambi i casi "Francesco" è un nome, ma nel secondo caso è stato necessario specificare di *che tipo* di dato si tratta (un nome); i termini che indicano il tipo di dato sono detti *metadati*.

Anche la bibliografia, come scienza degli indici, si occupa di metadati da quando esiste. Utilizzano metadati i *Pinakes* di Callimaco, i *loci communes* dei manoscritti medievali, le descrizioni bibliografiche di Konrad Gesner, il catalogo di Antonio Panizzi e tutti gli strumenti di mediazione fino allo standard ISBD. È invece relativamente recente in bibliografia l'utilizzo diffuso del *termine* metadati. Esso entra nel linguaggio biblioteconomico quando, con l'applicazione degli elaboratori ai processi di gestione dei dati bibliografici, si passa dal loro utilizzo implicito alla loro formalizzazione logica, allo scopo di creare modelli concettuali complessi sempre più adeguati alla rappresentazione dei fenomeni bibliografici. La definizione comune di metadati è meno chiara di quella proposta nel 2010 da Karen Coyle, secondo la quale i metadati sono "dati costruiti per risolvere un problema o per svolgere una funzione".³ Il collegamento dei metadati con una o più funzioni è molto utile nel contesto della bibliografia, perché permette di evidenziare come si è svolta, dal punto di vista dei metadati, l'evoluzione degli strumenti di descrizione delle risorse bibliografiche: dal catalogo manoscritto, al catalogo a stampa, al catalogo a schede a quello elettronico. Questa evoluzione ha segnato una costante crescita della *qualità* dei dati creati che, secondo la definizione dello standard ISO 9000:2005, si intende come "la totalità delle peculiarità e delle caratteristiche di un prodotto, processo o servizio che è in relazione alla capacità di soddisfare bisogni espliciti o impliciti".

Nel catalogo a libro i metadati erano di norma impliciti: in quel contesto tecnologico semplice, la funzione svolta dai metadati è quella base, ovvero rappresentare le risorse. Nel cata-

³ Karen Coyle, "Nuove direzioni in metadata [video]," 2010. <<http://www.archive.org/details/NuoveDirezioniInMetadata>>.

logo manoscritto e in quello a stampa, non c'è alcun bisogno di rendere espliciti i metadati. Questa necessità si presenta, invece, quando alla funzione fondamentale (la rappresentazione dei documenti) si aggiunge una nuova funzione: lo scambio dei dati bibliografici. Per svolgere questa funzione in modo economico è necessario e sufficiente scambiare e condividere e riutilizzare solo la descrizione delle entità di interesse: l'unità di descrizione della raccolta, rappresentata fino a quel momento dal catalogo nel suo complesso, dev'essere spezzata in più sottounità: ciò avviene nel passaggio dal catalogo manoscritto e a stampa al catalogo a schede. In questo caso, per esempio, occorre sostituire il trattino lungo che indicava la medesima intestazione con il dato esplicito nell'intestazione di ciascuna singola scheda. Se si osserva a posteriori il catalogo manoscritto, a stampa e a schede, appare evidente che una parte rilevante dei metadati è espressa dallo *stile di citazione*, ovvero "la soluzione fornita da un autore all'insieme dei problemi costituiti dalla scelta e dalla forma della formula citazionale, dall'organizzazione dei dati descrittivi, della fisionomia o presentazione finale della notizia bibliografica".⁴ La forma e l'ordine di presentazione dei dati costituiscono la modalità di registrazione dei metadati negli indici bibliografici in forma di libro e a schede. La particolare presentazione di un dato è un metadato importante perché consente al lettore di ricostruire le relazioni tra i vari elementi; per esempio, le relazioni del primo elemento (di norma l'autore, in maiuscolo o maiuscoletto) con il secondo elemento (di norma il titolo dell'opera, in corsivo) e di questi con gli elementi successivi (vedi figura 3).

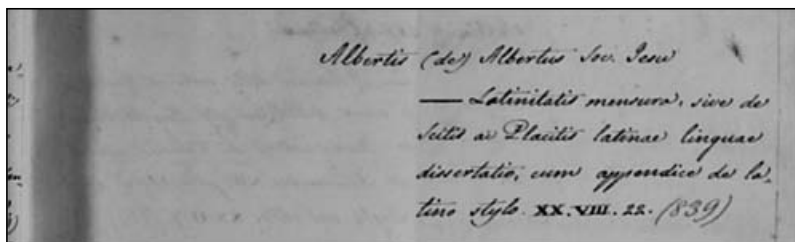


Figura 3 – Esempio di descrizione in un catalogo manoscritto

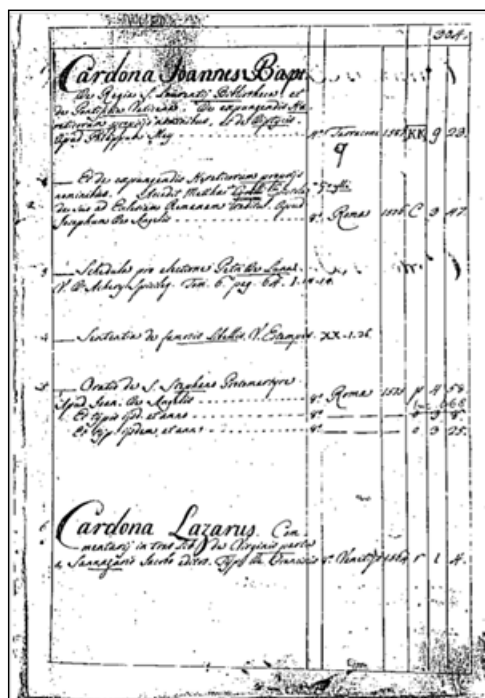
⁴ Marco Santoro, con la collaborazione di Gianfranco Crupi, *Lezioni di bibliografia*, Milano: Editrice Bibliografica, 2012, p. 178.

In questa prospettiva, la definizione di metadati proposta da Coyle è molto più suggestiva: in due descrizioni bibliografiche del tipo:

MILL, JOHN S.. Utilitarianism. London; 1863, 8vo. (1)

----- On liberty; 3d Ed. Boston; 1864, 16mo. (2)

il carattere tipografico e la sequenza ordinata dei dati sono lo strumento che svolge la funzione di indicare, per esempio, che l'es-



pressione “MILL, JOHN S.” si riferisce all’ autore e che le espressioni “Utilitarianism” e “On liberty” sono i rispettivi titoli. Sono caratteristiche che rispondono al bisogno, esplicito o implicito del lettore, di distinguere con maggiore facilità – e risparmio di tempo – il tipo di dati che costituiscono la rappresentazione. La struttura fornisce informazioni, ovvero dati sui dati, come nell’esempio in figura 4, tratto dal Catalogo per autori e titoli del fondo antico a stampa, redatto tra il 1756 e il 1786, della Biblioteca Angelica di Roma.⁵

Figura 4 – Dal Catalogo per autori e titoli del fondo antico a stampa, redatto tra il 1756 e il 1786, della Biblioteca Angelica di Roma

⁵ <http://cataloghistorici.bdi.sbn.it/dett_catalogo.php?IDCAT=181>.

Nelle due descrizioni bibliografiche (1) e (2) e in figura 4, il contesto fornisce i metadati. Ciò è possibile soltanto perché chi interpreta il contesto, cioè il lettore, è in grado di:

- 1) distinguere tra un tipo di dato e l'altro; per esempio, tra il titolo e il luogo o tra il titolo e l'edizione;
- 2) assegnare a ciascun dato il valore specifico, a prescindere dalla forma con la quale si presenta;
- 3) ricostruire esattamente le relazioni tra i dati e tra le entità che esse rappresentano.

Nei cataloghi a libro, anche la posizione relativa di una descrizione è un metadato: per esempio, la descrizione (2) si avvale di parte dei dati della descrizione (1) per economizzare la registrazione. Infatti, le linee che aprono la descrizione (2) e, in figura 4, le descrizioni bibliografiche relative a Giovanni Battista Cardona successive alla prima sostituiscono il nome dell'autore e indicano che l'autore non è cambiato dalla registrazione precedente alla registrazione successiva. Questa tecnica è possibile perché l'indice che informa sull'universo bibliografico costituisce un'unità indivisibile, atomica: il catalogo a libro. Ogni pagina dev'essere letta nel suo complesso e la legatura a volume garantisce l'ordinamento corretto delle informazioni presentate (ovvero le loro relazioni). È opportuno sottolineare infine che, in questa fase del catalogo manoscritto, a stampa e a schede, nella singola scheda esiste sovrapposizione tra modalità di registrazione e di presentazione dei dati: i dati sono presentati esattamente nel modo in cui sono stati registrati e ordine e forma di presentazione dei dati costituiscono i metadati necessari per la comprensione della scheda.

Nel passaggio tecnologico che porta dal catalogo a volume al catalogo a schede mobili il contesto cambia perché, dal momento che la posizione di ciascuna scheda può cambiare rispetto a quella che precede o segue, tutte le schede devono essere autonome dal punto informativo. Infatti, la tecnologia del catalogo a schede consentiva di svolgere una funzione aggiuntiva rispetto a quella base di rappresentazione degli oggetti bibliografici: l'interpolazione delle schede e la correzione e sostituzione di dati nel catalogo senza compromettere l'ordinamento (le relazioni) del catalogo stesso. La maggiore

granularità dei dati – che si può definire come il livello di risoluzione raggiunto da un sistema di archiviazione nella registrazione logica e fisica dei dati – garantita dalla nuova tecnologia (registrazione su schede) richiede però una compensazione nella creazione dei dati. Essa consiste in una ridondanza di dati all’inizio di ciascuna registrazione, la ripetizione del nome dell’autore in ogni singola scheda. Un vantaggio che ne deriva è che il catalogo può scomporsi più facilmente in sequenze di ordinamento diverse: sequenza con accesso per autore e titolo, sequenza con accesso per soggetto, sequenza con accesso per classificazione, etc., con aumento della quantità e qualità delle relazioni evidenziate. In questo modo, le descrizioni (1) e (2), essendo su schede diverse e “decontestualizzate”, dovranno apparire per esempio nella forma (3), (4) e (5):

(3) MILL, JOHN S.. Utilitarianism. London; 1863, 8vo.

(4) MILL, JOHN S.. On liberty; 3d Ed. Boston; 1864, 16mo.

(5) Utilitarismo

MILL, JOHN S.. Utilitarianism. London; 1863, 8vo.

In questo tipo di catalogo le descrizioni sono sufficienti a svolgere le funzioni richieste: la funzione tradizionale di rappresentazione delle risorse (funzione indicale) e la funzione nuova di suddivisione delle informazioni in unità minime interpolabili (ma con ridondanza nel riferimento al nome dell’autore).

La prima apparizione esplicita dei metadati avviene nel passaggio successivo, quando si cerca di svolgere la nuova funzione di “favorire lo scambio internazionale dei dati bibliografici” (ISBD(M), paragrafo 0.1.2.), ovvero:

- a) rendere possibile lo scambio di registrazioni provenienti da fonti diverse, in modo che le registrazioni prodotte in un paese possano essere facilmente accettate in cataloghi di biblioteca o in altri elenchi bibliografici di ogni paese;
- b) favorire l’interpretazione delle registrazioni oltre le barriere linguistiche, in modo che registrazioni prodotte per gli utenti di una lingua possano essere interpretate dagli utenti di altre lingue;

- c) favorire la conversione delle registrazioni bibliografiche in forma elettronica.

In particolare, l'esigenza di soddisfare le due nuove funzioni b) e c) – l'interpretazione delle registrazioni oltre le barriere linguistiche e la conversione delle registrazioni in forma elettronica – richiede lo sviluppo di appositi metadati, per indicare come devono essere separati i vari elementi dei dati (per esempio, luogo vs editore vs anno) e la funzione reciproca tra i dati (responsabilità principale vs altre responsabilità). I dati della (3) assumono la forma:

(6) Utilitarianism / by John Stuart Mill. – London : Parker, son, and Bourn, 1863. – 95 p. ; 22 cm

I metadati sono sempre in forma codificata, distinguibili mediante la sequenza ordinata degli elementi e la punteggiatura, che si specializza e diventa più complessa che in precedenza. Alla granularità dei dati d'autorità (i diversi cataloghi) e gestionali, si aggiunge un ulteriore livello di granularità, presente nella nuova organizzazione dei dati descrittivi – che si strutturano in aree funzionali (Area del titolo e della formulazione di responsabilità, Area dell'edizione, etc.), necessarie per svolgere le nuove funzioni (comprensione interlinguistica e conversione elettronica).

A ciò corrisponde una maggiore complessità dei metadati: nella forma (6) essi sono necessari a distinguere le aree tra loro e gli elementi all'interno di ciascuna area. Un effetto collaterale negativo di questo passaggio è che il linguaggio di registrazione dei dati si complica e ciò incide sulla rappresentazione dei dati, che inizia a essere meno trasparente per il lettore (in contraddizione con il secondo degli obiettivi prefissati).

Nel passaggio dal catalogo a schede a quello elettronico si aggiungono nuove funzioni: per esempio, per evitare la ridondanza nell'inserimento dei dati e per ottenere un maggiore controllo delle forme dei nomi, le intestazioni per autore vengono sostituite da un'unica registrazione d'autorità, posta in un archivio separato; per evitare la ripetizione dell'intera descrizione bibliografica, questa viene associata ai dati gestionali di una o più copie, registrati anch'essi in un archivio a parte. In questi passaggi

in cui aumenta la granularità dei cataloghi, cresce in proporzione la necessità di esplicitare i metadati, per consentire allo strumento tecnologico in uso di riconoscere e ricostruire correttamente l'unità di ciascuna descrizione bibliografica. Anche le relazioni tra i dati che sono parti di un'unica descrizione logica e sono registrati in archivi diversi devono essere esplicitate: il valore dei metadati espresso nella sintassi del linguaggio documentario (che rimane valida per la comprensione dei dati da parte del lettore) non è più sufficiente, perché tali relazioni devono essere gestite dalle macchine.

Si verifica quindi la distinzione funzionale tra dati relativi alla descrizione (cioè destinati alla rappresentazione sintetica di un'unità documentaria) e dati relativi all'accesso (cioè destinati a garantire l'individuazione e il recupero delle descrizioni pertinenti e rilevanti). L'obiettivo è aumentare la qualità dei dati e evitare ridondanze nella registrazione, oltre che favorire lo scambio dei dati. Dal punto di vista tecnologico, questa nuova funzione si è espressa con il passaggio da un archivio unico a più archivi: l'archivio delle registrazioni e gli archivi d'autorità, che registrano le forme ammesse e non ammesse dei nomi delle entità in base alle quali si può effettuare la ricerca (per esempio persone, enti, soggetti, classi).



Figura 5 – Visualizzazione di una registrazione in un OPAC

L'aumento della granularità del catalogo senza la distinzione tra *registrazione* dei dati e loro *rappresentazione* ha comportato una certa ridondanza dei dati. Per esempio, in figura 5, la relazione tra la descrizione e alcune registrazioni dell'archivio d'autorità (*Livius, Titus* e *Pellegrino, Vincenzo*) non è sufficiente a eliminare la ripetizione di questo dato all'interno della descrizione (" / Tito Livio" e " ; a cura di V. Pellegrino"). La ridondanza deriva dal catalogo a schede, in cui le registrazioni di edizioni collegabili al

medesimo autore richiedono la ripetizione del nome dell'autore in ogni intestazione.

Questo modello logico di metadati è rimasto invariato nella creazione del MARC e dei diversi formati da esso derivati. Anche nel MARC, per esempio, l'autore dev'essere ripetuto in più campi (proprio per soddisfare funzioni diverse). Il formato è orientato alla ricostruzione della corretta *visualizzazione* dei dati e la maggior parte dei dati sono registrati in forma testuale; per esempio, la sequenza delle aree descrittive in molti sistemi è una sequenza di testo 'naturale'.⁶

Un altro aspetto critico dei dati bibliografici creati oggi è la cattiva gestione degli identificativi delle entità registrate negli archivi d'autorità. In questi archivi le entità sono identificate mediante codici identificativi, ma i codici sono assegnati localmente (cioè in ogni sistema) e non sono sistematicamente trascritti nelle registrazioni destinate allo scambio; perciò di norma i software non li gestiscono correttamente e non se ne avvantaggiano.⁷ Il modello corrisponde – nel complesso – agli attuali cataloghi elettronici della realtà italiana e di gran parte di quella mondiale. Ci sono due tendenze di un certo interesse in questo quadro:

- 1) la sempre più netta separazione degli archivi d'autorità, al punto che nei cataloghi si ricorre all'uso di archivi d'autorità "esterni", ovvero creati e mantenuti al di fuori dell'istituzione che produce il catalogo: per esempio, a livello nazionale il *Nuovo Soggettario* della BNCF⁸ e a livello internazionale il VIAF (Virtual International Authority File)⁹ e le LCSH (Library of Congress Subject Headings);¹⁰
- 2) la necessità di qualificare le relazioni tra gli elementi di dati che, nel loro insieme, sono necessari per la rappresentazione completa di un'unica unità documentale. Per questo motivo, devono essere istituite e qualificate le relazioni tra i dati dell'archivio delle descrizioni e i dati dell'archivio

⁶ Library Linked Data Incubator Group, *Library Linked Data Incubator Group final report*, paragrafo 3.1.3.

⁷ Library Linked Data Incubator Group, *Library Linked Data Incubator Group final report*, paragrafo 3.1.3.

⁸ <<http://thes.bncf.firenze.sbn.it/>>.

⁹ <<http://viaf.org/>>.

¹⁰ <<http://id.loc.gov/authorities/subjects.html>>.

d'autorità e, se ne esiste uno distinto, con l'archivio dei dati di collezione. Per esempio, dev'essere istituita una relazione tra i dati della descrizione (6) e la forma corretta dell'intestazione relativa a John Stuart Mill nell'archivio di autorità e tale relazione dev'essere qualificata, per esempio, in SBN con il codice 1, che indica una responsabilità principale, ovvero una relazione di autore.

Molti cataloghi sono attualmente fermi a questo stato di cose per una ragione pratica: il formato di scambio dei dati, il MARC – una famiglia di standard diffuso in tutto il mondo nel quale sono espresse milioni di registrazioni – è stato progettato per riprodurre e scambiare in forma elettronica i dati bibliografici secondo il modello del catalogo a schede. Il MARC non è in grado di rappresentare un complesso di relazioni molto più articolato, perché non è la finalità per la quale è stato progettato in origine.

I cataloghi elettronici così costituiti sono tuttavia entrati in crisi per due ragioni fondamentali, collegate a una struttura logica insoddisfacente:

- 1) perché poco efficienti, rispetto a costi di produzioni troppo alti, rispetto alle tecnologie a disposizione;
- 2) perché poco efficaci rispetto alle aspettative dei lettori, abituati sempre di più a percorsi di ricerca basati sulla navigazione, ovvero sulla scoperta di relazioni tra entità.

3. L'evoluzione delle funzioni del catalogo: FRBR

Il modello logico efficace e suggestivo delineato nel rapporto FRBR ha riavviato la riflessione teorica sul catalogo e sulle regole di catalogazione, ha portato a individuare nuove funzioni per i dati bibliografici e ha evidenziato la necessità di nuove soluzioni tecnologiche per realizzarle.¹¹ Ci si è così resi conto che le regole in uso per

¹¹ Vedi: Elaine Svenonius, *The Intellectual Foundation of Information Organization*; Mauro Guerrini, *Verso nuovi principi e nuovi codici di catalogazione*, a cura di Carlo Bianchini, Milano: Sylvestre Bonnard, 2005; Carlo Bianchini, *Riflessioni sull'universo bibliografico. Funzioni, oggetti e modelli della catalogazione per autore e per titolo*, Milano: Sylvestre Bon-

creare i cataloghi erano obsolete quanto il formato per lo scambio dei dati;¹² nel 2001 Paul G. Weston poteva notare che “ad una così intensa attività di riflessione e progettazione a livello speculativo-concettuale non ha corrisposto, almeno fino a oggi, una altrettanto vigorosa attività di revisione dei codici normativi e degli standard per la registrazione elettronica dei dati bibliografici”.¹³

Sotto il profilo dell'efficacia dei cataloghi elettronici, il più importante documento prodotto dall'IFLA è stato il *Rapporto finale sui Requisiti Funzionali delle Registrizioni Bibliografiche*, noto come FRBR.¹⁴ In questo *Rapporto* l'efficienza dei cataloghi è stata esaminata nella prospettiva della reale funzione (requisito funzionale) svolta da ogni singolo dato bibliografico da registrare (attributi delle entità) in relazione a determinati obiettivi da raggiungere (funzioni utente). L'attenzione era rivolta a stabilire il numero minimo di dati da registrare per garantire all'utente del sistema la possibilità di raggiungere i propri obiettivi, ovvero garantire la qualità riducendo i costi. Le funzioni svolte dall'utente erano, quindi, il punto di partenza: ne sono state individuate quattro: trovare, identificare, selezionare e ottenere. I dati tradizionalmente presenti nelle descrizioni bibliografiche sono stati associati ad alcune entità, che costituiscono il focus dell'interesse dell'utente e che sono state riunite in tre classi: Gruppo 1 (*opera, espressione, manifestazione e item*), Gruppo 2 (*persona e ente*) e Gruppo 3 (*concetto, oggetto, evento e luogo*).

Il potere esplicativo e l'efficacia del modello interpretativo di FRBR ne hanno garantito l'enorme successo, che si è tradotto

nard, 2005; Carlo Bianchini, *Rassegna critica dei recenti sviluppi della catalogazione in ambito internazionale*, “Bollettino AIB”, vol. 49, n. 1 (gennaio 2009), p. 35-44.

¹² Roy Tennant, *MARC must die*, October 15, 2002, <<http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>>.

¹³ Paul Gabriele Weston, *Catalogazione bibliografica. Dal formato MARC a FRBR*, “Bollettino AIB”, 41, n. 3 (settembre 2001), p. 267-283.

¹⁴ IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records. Final Report*, München: K.G. Saur, 1998, <<http://www.ifla.org/files/cataloguing/frbr/frbr.pdf>>; traduzione italiana: IFLA Study Group on the Functional Requirements for Bibliographic Records, *Requisiti Funzionali per Record Bibliografici*, Roma: ICCU, 2000.

in un nuovo filone di studi e nella realizzazione, da parte dell'IFLA, di altri modelli ispirati allo schema di FRBR: *Functional Requirements for Authority Data* (FRAD) e *Functional Requirements for Subject Authority Data* (FRSAD).¹⁵

Le aspettative degli utenti rispetto alla possibilità di seguire percorsi di ricerca simili a quelli creati dal web erano state definite molto bene da Elaine Svenonius, in un'opera diventata un classico della letteratura biblioteconomica ma sfortunatamente pubblicata dopo FRBR. In quest'opera Svenonius notava che "alcuni utenti iniziano una ricerca d'informazioni sapendo esattamente che cosa vogliono; altri utenti però, non conoscono proprio, o non sanno esprimere, l'oggetto della loro ricerca, tuttavia sono in grado di riconoscerlo quando lo trovano. Questi utenti si aspettano di essere guidati e i sistemi bibliografici hanno soddisfatto questa aspettativa. Ne è un esempio la guida fornita da una classificazione usata per ordinare i libri sugli scaffali di una biblioteca".¹⁶ Ciò di cui hanno bisogno gli utenti che devono essere guidati nella ricerca è la *navigazione*: essa consiste nel seguire le connessioni esistenti tra le opere secondo relazioni di generalizzazione, associazione o aggregazione o nel trovare attributi correlati per equivalenza, associazione e gerarchia. La funzione navigare ha pertanto un valore informativo molto importante, che è entrato a fare parte di un'altra elaborazione teorica fondamentale prodotta dall'IFLA: i *Principi internazionali di catalogazione* (ICP).

¹⁵ IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), *Functional Requirements for Authority Data: a conceptual model. Final report, December 2008*, München: Saur, 2009; IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR), *Functional Requirements for Subject Authority Data (FRSAD): a conceptual model*, editors: Marcia Lei Zeng, Maja Žumer, Athena Salaba. Berlin: De Gruyter Saur, 2011.

¹⁶ Elaine Svenonius, *The intellectual foundation of information organization*, cit., p. 19. A parere di chi scrive, il migliore esempio di guida a scaffale mai realizzato da una classificazione è rappresentato dalla Colon Classification; vedi: Carlo Bianchini, *FRBR prima di FRBR. Il numero di libro nella Colon Classification*, "Jlis.it", vol. 1, no. 1 (gennaio/January 2010), p. 11-39 e Carlo Bianchini, *Organizzare la conoscenza con la sequenza di filiazione della Classificazione Colon di S. R. Ranganathan*, "Jlis.it", vol. 2, no. 2 (dicembre/December 2011), p. 1-21.

Secondo i *Principi internazionali di catalogazione* dell'IFLA del 2009,¹⁷ il catalogo è uno strumento efficiente ed efficace che consente all'utente di:

- trovare le risorse bibliografiche;
- identificare una risorsa bibliografica, cioè avere conferma che l'opera descritta corrisponda a quella cercata o poter distinguere fra due o più opere con caratteristiche simili;
- selezionare una risorsa che risponda al formato, al contenuto e al supporto desiderati;
- ottenere un item, una copia fisica o elettronica di una risorsa;
- navigare tra le informazioni.

Navigare, in ambito biblioteconomico, è una funzione che caratterizza l'OPAC e le banche dati.¹⁸ Se si considera la funzione di navigazione in relazione all'universo bibliografico, agli strumenti e alle tipologie di ricerca degli utenti, essa consente di contrapporre nettamente i cataloghi e i motori di ricerca.

I cataloghi – sul piano teorico¹⁹ – consentono di organizzare in modo efficace risorse con caratteristiche molto composite e consentono di distinguere tra opere, espressioni, manifestazioni e item – e più in generale tra le entità definite nei modelli della famiglia di FRBR. In questo modo consentono la navigazione e possono rispondere efficacemente ai bisogni delle ricerche consapevoli ma soprattutto di quelle *esplorative*. Al contrario i motori di ricerca sono in grado di fornire risposte alla ricerca consapevole, ma non sono per ora in grado di organizzare le informazioni raccolte nelle categorie di opera, espressione e manifestazione, perché l'approccio con il quale sono stati costruiti

¹⁷ IFLA. *IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its glossary*, by Barbara Tillett & Ana Lupe Crisán (eds.), München: Saur, 2009, <<http://www.ifla.org/publications/ifla-series-on-bibliographic-control-37>>.

¹⁸ IFLA Task Force on Guidelines for OPAC Displays, *Guidelines for Online Public Access Catalogue (OPAC) displays*, The Hague, Netherlands, 2005, paragrafo 7, <<http://archive.ifla.org/VII/s13/guide/opacguide03.pdf>>. Traduzione italiana parziale pubblicata in: Andrea Marchitelli, Giovanna Frigimelica, *OPAC*, Roma, Associazione italiana biblioteche, 2012.

¹⁹ Ibidem, paragrafo 5.2.

è basato sulla potenza di calcolo e non su un modello elaborato come quello dei cataloghi. Inoltre i motori di ricerca non sono in grado di comprendere il significato delle richieste che vengono poste né di identificare le relazioni semantiche tra i documenti del web.²⁰ L'ordine dei risultati per rilevanza (*ranking*) non consente la navigazione *tra* i risultati ed è del tutto inadeguato a soddisfare le esigenze delle ricerche esplorative.

4. L'effetto di FRBR sui modelli e sulle regole

Dalla consapevolezza della necessità di basare la realizzazione dei cataloghi su un modello logico più esplicito hanno preso avvio via i lavori di rinnovamento delle regole di catalogazione in molti paesi, ma le realizzazioni più concrete sono state italiane e angloamericane.²¹ Il primo codice pubblicato che ha recepito le novità introdotte da FRBR – anche se non completamente²² – è stato italiano: *Regole Italiane di Catalogazione*. REICAT,²³

²⁰ Carlo Bianchini, *Il falso in internet: autorevolezza del web, information literacy e futuro della professione (seconda parte)*, "AIB studi", vol. 54, n. 2/3 (maggio/dicembre 2014), p. 229.

²¹ Per un primo orientamento vedi: Barbara B. Tillett, *RDA (Resource Description and Access): lo sviluppo di un nuovo codice di catalogazione internazionale*, "Bibliotime. Rivista elettronica per le biblioteche", vol. 11, n. 1 (marzo 2007), <<http://www2.spbo.unibo.it/bibliotime/num-xi-1/tillett.htm>>; Barbara B. Tillett, *The Bibliographic universe and the new IFLA Cataloging Principles: lectio magistralis in library science ... = L'universo bibliografico e i nuovi Principi di Catalogazione dell'IFLA: lectio magistralis di biblioteconomia ...*. Fiesole (Firenze): Casalini Libri, 2008, <<http://digital.casalini.it/9788885297814>>; Alberto Petrucciani, *Ragioni e principi della revisione delle RICA: per un nuovo codice italiano di catalogazione*, "Bollettino AIB", 45, n. 2 (giugno 2005), p. 149-185; Giovanna Mazzola Merola, *Verso le nuove RICA*, "Bollettino AIB", vol. 45, n. 2 (giugno 2005), p. 139-148.

²² Isa De Pinedo, Alberto Petrucciani, *Un approccio all'applicazione del modello FRBR alle regole di catalogazione italiane: problemi e possibili soluzioni*, "Bollettino AIB", vol. 42, n. 3 (settembre 2002), p. 267-278.

²³ ICCU, *Regole italiane di catalogazione*. REICAT, a cura della Commissione permanente per la revisione delle Regole italiane di catalogazione, Roma: ICCU, 2009.

quello che ha accolto in forma integrale le entità del modello FRBR è stato RDA Resource Description and Access.²⁴

Le applicazioni normative derivate da FRBR, come REICAT (codice di regole) e RDA (linee guida), e il rapporto del Library Linked Data Incubator Group, confermano la necessità di adottare modelli di registrazione e di rappresentazione dei dati più articolati di quelli attuali: per raggiungere gli obiettivi della maggiore efficienza dei cataloghi e garantire agli utenti la funzione di navigazione è necessario che i metadati siano caratterizzati da:

- 1) maggiore granularità, ovvero un insieme maggiore di elementi di dati da riconoscere e da registrare;
- 2) maggiore complessità e precisione nella rappresentazione delle entità e delle relazioni, ovvero del *data model*, perché sono necessari più legami per collegare un numero maggiore di dati pertinenti alla stessa rappresentazione;
- 3) impiego di nuovi formati e nuovi strumenti tecnologici per la pubblicazione dei dati – le “5 stelle” descritte da Tim Berners-Lee²⁵ – che consentano la corretta interpretazione dei dati e il loro impiego da parte degli elaboratori (ma in ultima istanza da parte dei motori di ricerca), ovvero il ricorso al modello dei linked data, agli URI e a RDF.

REICAT e RDA sono intervenuti in questa direzione, aumentando il numero dei metadati, elaborando un modello di dati più granulare e rendendo possibile una maggiore complessità nella rappresentazione delle entità bibliografiche, secondo uno scenario tecnologico nuovo e articolato.²⁶ Le entità che si possono o si devono descrivere sono stabilite dalle regole catalografiche o dalle istruzioni che si applicano; nel caso di RDA, l'insieme delle entità che si possono descrivere è definito da appositi registri, che si

²⁴ Joint Steering Committee for Development of RDA, *RDA: Resource Description and Access*, 2013, <<http://www.rda-jsc.org/rda.html>>.

²⁵ Tim Berners-Lee, *Linked Data – Design Issues*, 2006, <<http://www.w3.org/DesignIssues/LinkedData.html>>.

²⁶ Carlo Bianchini, *Futuri scenari: RDA, REICAT e la granularità dei cataloghi*, “Bollettino AIB”, 50, n. 3 (settembre 2010), p. 219-238, <<http://bollettino.aib.it/article/view/5319/5082>>.

chiamano *RDA element sets*.²⁷ In RDA sono previsti 14 elementi per il registro delle *FRBR entities for RDA* (che corrispondono a opera, espressione, manifestazione, item, persona, famiglia, ente, concetto, etc.), 469 elementi per il *RDA Group 1*, 62 elementi per il *RDA Group 2* e 19 elementi per il *RDA Group 3*. Nel nuovo scenario, è necessario registrare con cura le relazioni che intercorrono tra le entità, sia collegando le entità sia qualificando il valore della relazione in base a un insieme predefinito nei *vocabularies*.

Il codice e le linee guida segnano un cambiamento positivo e profondo per la teoria della catalogazione; RDA è coraggiosamente andato oltre il presente²⁸ e interpreta gli sforzi dell'IFLA nel creare nuovi principi (ICP).²⁹ Il modello FRBR ha influito molto anche su ISBD, lo standard descrittivo prodotto dall'IFLA: esso infatti è alla base di un processo di rinnovamento durato diversi anni e che ha portato alla pubblicazione di un'edizione consolidata di ISBD.³⁰ Rimangono aperti diversi problemi, che l'IFLA deve risolvere se vuole garantire continuità allo standard che ha riscosso un notevole successo in passato: chiarire il significato del concetto centrale di risorsa e le sue modalità di rappresentazione,³¹ realizzare i passaggi necessari per rendere

²⁷ Joint Steering Committee for Development of RDA, *The registry! RDA vocabularies*, 2012, <<http://rdvocab.info/>>.

²⁸ Vedi: Barbara B. Tillett, *Mantenere il ruolo delle biblioteche nel web semantico tramite RDA: Resource Description and Access*, "Jlis.it", vol. 2, no. 2 (luglio/July 2011), p. 1-19; Maria Grazia Iorio, Tiziana Possemato, *RDA: la catalogazione ai tempi del web*, "Biblioteche oggi", vol. 29, n. 9 (novembre 2011), p. 12-16.

²⁹ Giuliano Genetasio, *The International Cataloguing Principles and their future*, "Jlis.it", vol. 3, no. 1 (gennaio/January 2009), p. 1-17; Mauro Guerrini, *I Principi internazionali di catalogazione (ICP)*, "Biblioteche oggi", vol. 29, n. 9 (novembre 2011), p. 5-11.

³⁰ IFLA. ISBD Review Group, IFLA Section on Cataloguing, *ISBD International Standard Bibliographic Description. Consolidated Edition. Recommended by the ISBD Review Group. Approved by the Standing Committee of the IFLA Cataloguing Section*, Berlin ; Munchen: De Gruyter Saur, 2011, <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2012/ISBD_2012_trad_it_online.pdf>.

³¹ Vedi: Carlo Bianchini, Mirna Willer, *ISBD resource and its description in the context of the semantic web*, "Cataloging & Classification Quarterly", vol. 52, no. 8 (2014), p. 869-887.

ISBD pienamente adatto al web semantico³² e infine delineare i rapporti tra ISBD, il formato UNIMARC e RDA.³³

5. RDA e web semantico

La novità più recente e importante per l'integrazione dei dati delle risorse bibliografiche nel web semantico è rappresentata da RDA. Gli aspetti che rendono interessanti le linee guida nel nuovo contesto del circuito dell'informazione bibliografica sono numerosi; RDA è:

- basato su principi internazionali condivisi;
- rivolto a una platea internazionale (auspicabilmente mondiale);
- destinato alle agenzie che creano i dati, nonché all'uso e all'integrazione dei dati di tutte le istituzioni culturali (biblioteche, archivi e musei, etc.);
- applicabile a qualsiasi tipo di risorsa;
- pensato per l'ambiente digitale e il web;
- progettato per l'efficienza e la facilità d'uso;
- utilizzabile in contesti tecnologici diversi.

RDA è basato su principi e modelli elaborati e condivisi a livello internazionale: il richiamo a FRBR e ai modelli derivati è esplicito ed è posto all'inizio del testo delle linee guida: "Un elemento chiave nella progettazione di RDA è il suo allineamento con i modelli concettuali per i dati bibliografici e d'autorità sviluppati dall'International Federation of Library Associations

³² Elena Escolano Rodríguez, *ISBD en la web semántica : lectio magistralis en Biblioteconomía : Florencia, Italia, Universidad de Florencia, 6 de marzo de 2012* = *ISBD nel web semantico : lectio magistralis in Biblioteconomia : Firenze, Italia, Università di Firenze, 6 marzo 2012*. Fiesole (Firenze): Casalini Libri, 2012; Gordon Dunsire, *The Role of ISBD in the Linked Data Environment*, "Cataloging & Classification Quarterly", vol. 52, no. 8 (2014), p. 855-868.

³³ Mirna Willer, Gordon Dunsire, *ISBD, the UNIMARC Bibliographic Format, and RDA: Interoperability Issues in Namespaces and the Linked Data Environment*, "Cataloging & Classification Quarterly", vol. 52, no. 8 (2014), p. 888-913.

and Institutions”, ovvero FRBR e FRAD.³⁴ L’allineamento concettuale tra il modello e le linee guida è pieno, senza riserve.³⁵

Le entità descritte in RDA sono tutte derivate da FRBR e da FRAD, anche se per le entità *concetto*, *oggetto* ed *evento* non sono state ancora sviluppate le linee guida.³⁶ RDA include inoltre un riferimento preciso a ICP: “Lo *Statement of International Cataloguing Principles* dell’IFLA informa i principi di catalogazione adottati in RDA”.³⁷

FRBR, FRAD e ICP forniscono a RDA la struttura di base e i presupposti per sviluppare alcune sue caratteristiche più innovative:

- a) l’estensione e la flessibilità necessarie a coprire tutti i tipi di contenuto intellettuale e di media;
- b) la modularità indispensabile per adattarsi a nuove future risorse che potranno diffondersi in futuro;
- c) l’adattabilità richiesta per creare dati validi in un’ampia gamma di ambienti “tecno-logici”, ovvero per essere applicabili nei cataloghi del futuro e perfino in quelli di oggi.³⁸

Mai, prima di RDA, un insieme di linee guida per la descrizione e l’accesso alle risorse si era proposto con uguale determinazione come standard (questo è il termine di autodefinizione di RDA) di rilevanza “globale”: a partire dal titolo, dal quale è scomparso ogni riferimento geografico e culturale all’area anglo-americana che aveva caratterizzato AACR2, delle quali RDA è il successore. L’esistenza di un Gruppo Europeo di interesse per RDA (EURIG, European RDA Interest Group)³⁹ e la presenza della Deutsche Nationalbibliothek nel Committee of Principals e nel Joint Steering Committee for Development of RDA, rispettivamente gli organi di governo politico e tecnico, sono più che

³⁴ Joint Steering Committee for Development of RDA, *RDA: Resource Description and Access*, paragrafo 0.3.

³⁵ Ibidem, paragrafo 0.3.2 e 0.3.3.

³⁶ Ibidem, paragrafo 0.3.3.

³⁷ Ibidem, paragrafo 0.4.1.

³⁸ Carlo Bianchini, *Futuri scenari: RDA, REICAT e la granularità dei cataloghi*, cit., p. 225.

³⁹ <<http://www.slainte.org.uk/eurig/>>.

indizi dell'interesse che RDA suscita al di fuori dell'area di influenza tradizionale delle AACR. L'interesse è ampio e ciascuna realtà nazionale può contribuire apportando le proprie idee e conoscenze, come dimostra l'esempio della Biblioteca Nacional de España. Essa riconosce che RDA pone l'accento su due obiettivi fondamentali e condivisi, come la trasformazione del catalogo nel senso di FRBR e lo sviluppo dei linked data, ma evidenzia anche che per adottare RDA ci sono difficoltà concrete legate alla poca stabilità del testo di RDA, al fatto che l'attuale formato in uso (MARC21) non è abbastanza flessibile, a questioni legate ai dati già prodotti e ai costi da sostenere.⁴⁰

RDA si propone come standard di descrizione e accesso a qualsiasi tipo di risorsa: è una raccolta di linee guida per la registrazione di dati per qualsiasi oggetto conservato e accessibile in biblioteca (manoscritti, libri, periodici, musica, carte geografiche, film, ebook, etc.), in archivio (documenti istituzionali, carte personali e familiari, documentazione aziendale, lettere etc.), in museo (opere d'arte, abiti di scena, reperti e oggetti naturali, aeromobili e macchine spaziali, modellini, etc.) e delle multiformi risorse prodotte e disseminate usando le tecnologie digitali. Sono numerose le voci del glossario che segnalano la decisa apertura agli oggetti culturali non strettamente bibliografici: dipinto (painting), stemma (coat of arms), moneta (coin), pittura a olio (oil paint), collage (collage), diorama (diorama), scheda didattica (flash card), medaglia (medal), radiografia (radiography), immagine di telerilevamento (remote sensing image), etc. Per favorire l'uso delle linee guida anche all'interno della comunità archivistica, per esempio, RDA ha accolto l'entità Famiglia (inizialmente non prevista in FRBR) e ha definito elementi come la *data di produzione* (che registra dati cronologici relativi a risorse non pubblicate). Va precisato, tuttavia, che una maggiore quantità di esempi dedicati a questi oggetti avrebbe certamente favorito questo processo di apertura. Su alcuni punti – come le linee guida sull'identificazione, descrizione e accesso delle manifestazioni (RDA, Capitoli 2, 3 e 4), o le appendici – RDA è certamente carente e poco innovativo rispetto ad AACR2. Anche l'aspirazione

⁴⁰ <http://www.bne.es/webdocs/Prensa/Noticias/2014/1218-Comunicado_RDA-ingles.pdf>.

a rivolgersi a una platea internazionale è spesso mortificata da scelte determinate dal contesto culturale angloamericano.

Ci sono due aspetti di esplicita rottura rispetto agli strumenti precedenti: la struttura dello standard e la granularità dei dati. RDA non prevede una parte dedicata alla descrizione bibliografica (sviluppata dai codici precedenti). RDA è strutturato in due sole parti, pienamente corrispondenti alle entità e alle relazioni definite nel modello FRBR; le sezioni 1-4 sono dedicate alla registrazione degli *attributi* delle entità; le sezioni 5-10 alla registrazione delle *relazioni* tra entità.

Identificare e collegare un'entità sono, quindi, i due obiettivi fondamentali di RDA.

Le sezioni 1-4, dedicate agli attributi delle entità, indicano quali dati sono necessari per identificare tutte le entità, tra cui *anche* manifestazione e item, ovvero l'oggetto della descrizione bibliografica nei codici precedenti. Se si esaminano le linee guida nella parte relativa all'identificazione di manifestazione e item, si nota che il cambiamento è soprattutto di natura logica. Come si è visto nel paragrafo 1, la descrizione di un oggetto non comporta necessariamente la sua identificazione. Solo se il numero di caratteristiche incluse nella descrizione è sufficientemente alto, la descrizione è anche identificativa; tuttavia il numero delle caratteristiche necessarie aumenta in relazione alle dimensioni dell'insieme di risorse che si descrivono. RDA pone l'accento sul fatto che la semplice descrizione di una manifestazione o un item (prevista nei codici precedenti) non è sempre sufficiente a identificare quella risorsa e che invece l'obiettivo finale dev'essere proprio l'identificazione.

Il cambiamento non è, quindi, segnato da un'assenza (quella di una parte del codice dedicato alla descrizione bibliografica della pubblicazione/manifestazione, intesa come registrazione delle caratteristiche rilevanti della risorsa) quanto dalla circostanza che lo scopo è diventato l'identificazione e che gli oggetti del processo d'identificazione si sono moltiplicati, per includere qualsiasi entità del modello.

Identificare implica la registrazione degli attributi di un'entità, con un procedimento assimilabile a quello della creazione di un authority record per quell'entità. Per questo motivo, le linee guida rendono sistematica l'identificazione delle entità descritte

in FRBR: persone, famiglie, enti, opere, espressioni, manifestazioni e perfino item. Questo approccio, che richiede l'identificazione sistematica di tutte le entità coinvolte, aumenta il numero di authority file da creare e gestire.

A questo punto del processo le entità sono identificate, ma non sono ancora state chiarite le relazioni che intercorrono tra di esse; per esempio, potremmo avere i dati che identificano una persona e i dati che identificano un'opera, ma non sapere che tra di essi c'è un collegamento, una relazione. È questo il motivo per cui il secondo obiettivo di RDA è collegare le entità sulla base di relazioni concettuali e funzionali. Le linee guida dedicano maggiore spazio alla registrazione delle relazioni rispetto ai codici precedenti; è infatti un passaggio indispensabile per dare sostanza alla funzione *navigare* tra le entità, tra dati di tipo e origine diversa, funzione concepita da Elaine Svenonius e ripresa da ICP.⁴¹

Nella prospettiva del web semantico, una caratteristica fondamentale delle relazioni di RDA è che sono qualificate. Le appendici I, J, K e L sono dedicate ai designatori di relazione, ovvero termini controllati che forniscono informazioni più specifiche sulla natura della relazione che sussiste tra due entità. Per riprendere l'esempio citato, la relazione tra un'entità persona e un'entità opera dev'essere anche qualificata, specificando il ruolo svolto dalla persona: la persona è un *creatore* dell'opera; inoltre è possibile anche qualificare in modo più preciso il ruolo, con termini più specifici di creatore, come *autore*, *artista*, *regista*, *progettista*, *compositore*, *inventore*, *programmatore*, etc. (RDA, paragrafo I.2).

Una importante innovazione di RDA è rappresentata dalla diversità di modelli tecnologici con cui può essere utilizzato e dal vantaggio che questa flessibilità comporta: RDA è progettato per le tecnologie attuali, ma è aperto a quelle del futuro, in particolare a quelle necessarie per popolare il web semantico con dati bibliografici di qualità.⁴² In RDA le relazioni sono qualificate,

⁴¹ Carlo Bianchini, Mauro Guerrini, *RDA: Resource Description and Access. Il nuovo standard per la metadattazione e la scoperta delle risorse nell'era digitale*, "Biblioteche oggi", vol. 32, n. 4 (maggio 2014), p. 47.

⁴² Carlo Bianchini, *Dagli OPAC ai library linked data: come cambiano le risposte ai bisogni degli utenti*, "AIB studi", vol. 52, n. 3 (settembre/dicembre 2012), p. 303-323.

proprio come richiesto da RDF, il modello di dati orientato al web semantico. L'approccio di RDA nella registrazione dei dati è granulare; l'attenzione si sposta – in linea con la parabola compiuta nel passaggio da FRBR a FRAD – dal *record bibliografico* ai singoli *dati* di cui esso è costituito. Secondo uno sviluppo che procede nella stessa direzione intrapresa quando nel catalogo elettronico è iniziata la gestione degli authority file per persone, enti, soggetti e classi, la registrazione di tutti i dati bibliografici relativi a qualsiasi entità di RDA – ma più in generale del modello FRBR – deve segnare il passaggio “dalla gestione di record alla gestione di entità”.⁴³

RDA è utilizzabile in contesti tecnologici diversi perché è, come si definisce, uno *standard di contenuto*:⁴⁴ le caratteristiche elencate hanno consentito a RDA di diventare estremamente versatile in termini di allargamento dei destinatari-utilizzatori delle linee guida e di applicabilità nei contesti più vari. RDA si definisce, quindi, come uno standard di contenuto: “ciò significa che si applica per stabilire qual è l'informazione necessaria per descrivere una risorsa, cioè per stabilire il *contenuto* di una descrizione. RDA svincola le decisioni sul contenuto dei dati dalla tecnologia di registrazione dei dati (cartacea o elettronica), dal formato di registrazione dei dati (per esempio, MARC21 o UNIMARC o, in futuro, BIBFRAME) e da altri aspetti della descrizione che in precedenza, e in altri codici, sono strettamente connessi con la descrizione stessa: per esempio, la modalità di presentazione delle informazioni in aree prestabilite, l'ordine degli elementi, la punteggiatura prescritta, etc. Laddove una descrizione prevede di indicare che la risorsa oggetto di descrizione è un audiovisivo, RDA vincola il catalogatore a fornire questa informazione, ma non obbliga a mostrare l'informazione in una forma particolare”.⁴⁵

⁴³ Daniel Van Spanje, *Rise, fall of the cataloguer's empire: a changing landscape*, relazione presentata al convegno: *Faster, Smarter, Richer. Reshaping the library catalogue*, Roma (Italy), 27-28 February 2014, diapositiva n. 36, <<http://eprints.rclis.org/22835/>>.

⁴⁴ Barbara B. Tillett, *Prefazione*, in Carlo Bianchini, Mauro Guerrini, *RDA: Resource Description and Access*, cit., p. 13.

⁴⁵ Carlo Bianchini, Mauro Guerrini, *Introduzione a RDA*, cit., p. 68.

Queste caratteristiche hanno suggerito inoltre l'introduzione teorica e pratica di una netta linea di separazione tra le istruzioni che si occupano dei contenuti dei dati e quelle che ne definiscono la presentazione. In passato, per esempio nei cataloghi manoscritti o nei cataloghi a schede, l'informazione (ogni singolo dato) doveva essere registrato esattamente come sarebbe stato visto e letto dall'utente. Ancora nel *Classified catalogue code* di Ranganathan ci sono norme sulla grafia, sul carattere e sull'uso della matita per la registrazione di certi elementi dei dati descrittivi.⁴⁶ RDA sceglie deliberatamente di essere un insieme di linee guida che si occupa solo del contenuto dei dati, allo scopo di consentire modalità di registrazione dei dati applicabili indipendentemente da qualsiasi struttura o sintassi si desideri o si possa adottare per la loro archiviazione, il loro recupero e la loro visualizzazione.

L'obiettivo primario delle linee guida è stabilire come creare i dati necessari a identificare e fornire accesso alle risorse; quindi RDA non si propone di specificare come i dati devono essere visualizzati né di stabilire come devono essere codificati nei sistemi informatici. Per esempio, supponendo di avere una descrizione bibliografica presentata con il formato a bandiera (nel quale a sinistra si trova il nome del campo, per esempio "Titolo:" e a destra il valore relativo "I promessi sposi"), RDA stabilisce il valore del dato a destra, mentre non si occupa di stabilire se verrà visualizzato in un formato a bandiera, ISBD o altro formato di rappresentazione, né di come o dove quel dato verrà archiviato e recuperato dal sistema informatico, meccanico o manuale in uso presso una data agenzia bibliografica. Sulle ragioni della scelta, il JSC ha deciso di lasciare liberi gli utilizzatori "non professionali" di RDA di adottare gli strumenti a loro più confacenti e ha individuato nella definizione del contenuto dei metadati il *core business* dei bibliotecari.

La distinzione tra contenuto dei dati e loro forma di memorizzazione e presentazione ha importanti conseguenze teoriche e pratiche, sulla tradizione descrittiva basata sull'ISBD. Prima di tutto ISBD deve chiarire che cosa si intende per risorsa nel suo

⁴⁶ Shiyali Ramamrita Ranganathan, *Classified catalogue code with additional rules for dictionary catalogue Code*, 5th ed., Bombay: Asia Publishing House, 1964, p. 98-99.

contesto (per esempio la manifestazione, com'è stato tradizionalmente, o qualsiasi entità, come richiesto da FRBR) e in secondo luogo deve distinguere tra dati descrittivi basati su un approccio informativo (che mirano a esprimere il significato dei dati in una forma standard) o su un approccio iconico (cioè che mirano a riprodurre l'autopresentazione della risorsa).⁴⁷ La distinzione è fondamentale soprattutto nell'ottica dei linked data, in cui è il contesto d'uso dei dati che consente di stabilire se, in risposta a una determinata richiesta, sia opportuno restituire dati descrittivi di tipo informativo o iconico.

Sul piano pratico, la distinzione tra contenuto e forma dei dati è il motivo per cui l'ISBD, anche nell'ultima versione consolidata, è stata relegata in un'appendice (appendice D) di RDA e non fa, quindi, parte integrante del nuovo codice internazionale di catalogazione.⁴⁸ Al momento attuale, il ruolo e il futuro dello standard ISBD non sembrano chiari: il concetto di risorsa in ISBD si è modificato, durante la revisione avvenuta sotto la spinta di FRBR e nel processo di consolidamento dalle edizioni dedicate a specifici formati e forme bibliografiche al testo ISBD consolidato, ma la sua trasformazione in un concetto adatto a essere integrato nel web semantico non è ancora compiuta. Inoltre rimangono ancora aperti, per il momento, alcuni problemi nel processo di rappresentazione di ISBD e UNIMARC nel modello RDF, la base per il web semantico e i linked data.⁴⁹ A questo proposito, è utile ricordare che la Library of Congress e OCLC stanno collaborando per lo sviluppo di BIBFRAME,⁵⁰ un pro-

⁴⁷ Carlo Bianchini, Mirna Willer, *ISBD resource and its description in the context of the semantic web*, cit., p. 12-16.

⁴⁸ Carlo Bianchini, Mauro Guerrini, *Quis custodiet ipsos custodies? Osservazioni sulle relazioni fra FRBR, ICP, ISBD e RDA*, "Il Bibliotecario", vol. 3. n. 1 (gennaio-aprile 2009), p. 59-85, <<http://eprints.unifi.it/archive/00001743/>>.

⁴⁹ *ISBD Resource and its description in the context of the semantic web*, cit., p. 4-7; Willer, Dunsire, *ISBD, the UNIMARC Bibliographic Format, and RDA: interoperability issues in namespaces and the linked data environment*; Lynne C. Howarth, *ISBD as bibliographic content standard: interweaving threads, contemplating a future*, "Cataloging & Classification Quarterly", vol. 52, no. 8 (2014), p. 982-999.

⁵⁰ <<http://www.loc.gov/bibframe/>>.

getto destinato alla creazione di un nuovo formato di scambio bibliografico.⁵¹

La caratteristica che rende RDA immediatamente applicabile al web semantico è l'alto livello di separazione tra dati di contenuto e tutti gli altri dati necessari alla loro registrazione in base alla tecnologia adottata per la realizzazione dell'indice bibliografico, anche se è un processo ancora lontano dall'essere completato.

6. Nuovi dati per nuove funzioni. I library linked data

Il rapporto del gruppo formato dal W3C è probabilmente destinato ad avere sul mondo delle biblioteche un impatto altrettanto ampio, se non maggiore, di quanto non abbia avuto fino a oggi il rapporto FRBR. Esso afferma che per essere utilizzabili nel web, i dati bibliografici devono modificarsi per svolgere nuove funzioni: essere leggibili, modulari e riutilizzabili anche in ambienti non bibliografico-documentari, cioè nel web aperto. Come si è visto nel paragrafo 2, le trasformazioni tecnologiche del catalogo e dei suoi metadati sono state sempre determinate dalla necessità di svolgere nuove funzioni. Nel tempo si è lentamente e profondamente modificata la struttura del catalogo e l'organizzazione dei metadati: ogni aumento nel numero delle funzioni ha richiesto l'adozione di nuovi modelli logici, nuovi strumenti e nuovi formati di dati.⁵²

Come si è visto nel resto di questo volume, l'espressione linked data (dati connessi) si usa per indicare l'insieme delle buone pratiche per pubblicare, condividere e collegare singoli dati, informazioni e conoscenze sul web semantico usando gli URI e RDF.

⁵¹ Library of Congress, *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*, Washington, D.C.: Library of Congress, 2012, <<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>.

⁵² Karen Coyle, *Library Data in a Modern Context*, "ALA Techsource", vol. 46, no. 1 (2010), p. 5-7, <<http://alatechsource.metapress.com/content/p3022442071g7655/fulltext.pdf>>; Coyle, *Nuove direzioni in metadata [video]*".

L'URI è uno strumento per identificare una risorsa nel web. RDF indica il modello dei dati per i linked data. L'essenza di RDF è descrivere le relazioni tra gli oggetti sotto forma di triple (terne che combinano soggetti, predicati e oggetti), cioè rappresentare concetti (dati) molto più semplici di quanto non siano i contenuti di una pagina web, di una descrizione bibliografica o anche di una sua piccola parte.

Per creare un web basato sui dati invece che sui documenti è necessario un formato che consenta di creare legami tra semplici dati: RDF crea relazioni tra oggetti, tipicamente sotto forma di triple, rappresentate di solito mediante un *grafo*. La visualizzazione a grafo è il modello mentale più semplice possibile per l'RDF.⁵³ La struttura e il funzionamento di RDF sono descritti in questo volume al paragrafo 2.2.

Un'altra importante caratteristica da evidenziare è che nei linked data alcune triple che si ottengono nel processo di pubblicazione (*serializzazione*; vedi paragrafo 2.3.4) sono informazioni che fanno riferimento agli *element set* o ai *vocabularies*, che sono descritti una volta per tutte nel dominio relativo (cfr. cap. 3).

L'uso di un modello basato su strumenti semplici come URI e RDF consente di mescolare, visualizzare, condividere e, soprattutto, *inferire* dati (vedi paragrafo 1.5). A proposito di quest'ultimo processo, una caratteristica essenziale per comprendere la struttura del web semantico è che il processo di inferenza è possibile anche quando i dati strutturati e semi strutturati sono resi disponibili da applicazioni e agenzie molto diverse tra loro. Dal punto di vista dei dati di natura bibliografica, è importante sottolineare che un'agenzia che produce dati può creare triple RDF relative alle proprie raccolte con propri vocabolari e element set e che ma esse, una volta pubblicate, possono interagire con triple prodotte all'esterno del suo sistema informativo. Il modello RDF consente che dati di provenienza eterogenea si connettano e interagiscano tra loro, perché qualsiasi dato pubblicato con RDF contiene un'auto-descrizione (cioè è auto-consistente).

Fintanto che i dati sono registrati con formati proprietari e si

⁵³ Vedi, per esempio, la figura 8 di Mauro Guerrini, Tiziana Possemato, *Linked data: un nuovo alfabeto del web semantico*, "Biblioteche oggi", vol. 30, n. 3 (aprile 2012), p. 7-15.

trovano in archivi separati, come avviene per esempio con i formati della famiglia MARC, non sono visibili gli uni agli altri né collegabili tra loro. Se registrati come linked data invece, nel momento in cui un motore di ricerca effettua una ricerca nel web semantico, individua e recupera tutte le relazioni esistenti con uno o più dati che corrispondono a un certo URI, proprio come oggi cerca tutte le ricorrenze di una stringa di testo nelle pagine del web.

Entità e relazioni devono essere identificate in funzione almeno di un dato contesto; in effetti, la vera sfida dei linked data – secondo quanto fissato dalla quarta regola prevista da Tim Berners Lee (cfr. paragrafo 1.3)⁵⁴ – è che entità e relazioni siano identificate e collegate in *più* contesti. Ciò significa che nei linked data, oltre a creare i dati e i collegamenti necessari per descrivere in modo coerente un universo di oggetti, è necessario collegare quell'universo di dati ad altri universi di dati, per rendere possibile la navigazione attraverso il web semantico da parte dell'utente (vedi *interlinking*, paragrafo 2.5.1).

7. Condivisione, modularità e riutilizzo dei linked data bibliografici

Nel web costituito dai linked data, non esiste più la distinzione tra dati “bibliografici” e dati “di altra natura”; esistono certamente dati creati da agenzie che hanno finalità bibliografiche e *library linked data*, cioè dati pubblicati come linked open data e il cui contenuto abbia rilevanza bibliografica, ovvero possa essere utilizzato per svolgere le funzioni tipiche dei sistemi bibliografici. Nel web dei linked data esistono solo dati: condivisibili, modulari e riutilizzabili.⁵⁵ Con un'efficace immagine, Antonella De Robbio evidenzia che i dati bibliografici nel web semantico dovrebbero costituire la *Open biblio* (o *Open bibliographic data*), uno spazio intermedio tra i dati grezzi messi a disposizione liberamente dalla comunità scientifica e quelli messi a disposizione dalle istituzioni pubbliche: questo spazio intermedio “comprende i cataloghi delle biblioteche (OPAC), i database bibliografici, le basi di conoscenza, i tesauro e gli schemi di classificazione, gli

⁵⁴ Berners-Lee, *Linked Data - Design Issues*, cit.

⁵⁵ Ibidem.

archivi aperti e gli archivi di materiale didattico e-learning”.⁵⁶ La condivisione dei dati è un valore acquisito in ambito bibliotecario e a lungo perseguito dall’IFLA.

Nei linked data non ci sono problemi di obsolescenza dei formati o di conversioni tra formati: le relazioni tra dati e metadati sono registrate direttamente nei dati e, come avviene per RDA come standard di contenuto, si mantiene sempre la distinzione tra metadati relativi al valore semantico, al formato e alla sintassi. Ciò implica che i linked data conservano pienamente il loro significato anche quando cambiano formato.⁵⁷

Un aspetto fondamentale della condivisione dei dati, che dovrebbe consentire di superare l’immobilità che ha caratterizzato fino a oggi il formato dei dati bibliografici (e di favorire l’integrazione di dati prodotti da altre istituzioni culturali come archivi e musei), è che non sarà necessario ricreare o riconvertire l’enorme massa di dati creati finora, né ricreare il web in modo diverso: questo problema ha costituito uno dei maggiori ostacoli all’evoluzione dei modelli di dati dei cataloghi. Nel web dei dati sarà sufficiente predisporre i dati esistenti dotandoli dei metadati necessari per il web.⁵⁸ Per questo motivo, sottolineano Gordon Dunsire e Mirna Willer, “è estremamente importante sviluppare una rappresentazione RDF dell’UNIMARC come schema della struttura dei metadati allo scopo di estrarre il contenuto delle registrazioni in forma di linked data per il web semantico [...]”. Questa rappresentazione RDF sarebbe utile anche per aggiornare l’allineamento tra UNIMARC e gli altri standard che sono stati rivisti considerevolmente in anni recenti”.⁵⁹

Con i linked data non esiste un contributo così piccolo da non essere utilizzabile o vantaggioso. La modularità è un’altra caratteristica fondamentale per comprendere le potenzialità dei linked data bibliografici. Nel web dei dati, le relazioni tra i

⁵⁶ Antonella De Robbio, *Forme e gradi di apertura dei dati*, cit., p. 11.

⁵⁷ Library Linked Data Incubator Group, *Library Linked Data Incubator Group final report*, paragrafo 2.3.

⁵⁸ Ibidem, paragrafo 2.1.

⁵⁹ Gordon Dunsire, Mirna Willer, *Standard library metadata models and structures for the semantic web*, “Library Hi Tech News”, vol. 28, no. 3 (2011), p. 6; Gordon Dunsire, Mirna Willer, *UNIMARC and Linked Data*, “IFLA Journal”, vol. 37, no. 4 (December 2011), p. 314-326.

dati non sono sempre e soltanto riferibili alle entità definite nel modello FRBR o dell'universo bibliografico tradizionale delle biblioteche: si creano collegamenti tra dati prodotti all'interno della biblioteca e informazioni prodotte all'esterno e indipendentemente, da agenzie di natura non bibliografica. Modularità significa uso effettivo e immediata applicazione di dati anche in parte prodotti da altre agenzie (per esempio, i dati di pubblicazione forniti dall'editore, come l'ISBN, la copertina, la presentazione editoriale, etc.) e, viceversa, l'impiego diretto, da parte di altre agenzie, di dati bibliografici di alta qualità prodotti in ambito bibliotecario. In questo contesto, la diffusione dei dati bibliografici di qualità su nomi e soggetti prodotti e posseduti dalle biblioteche e la creazione delle relazioni tra dati bibliografici d'autorità e altri dati come *Wikipedia*, GeoNames, MusicBrainz, etc. è fondamentale per il web.⁶⁰

Secondo alcuni autori, nel web dei dati la percentuale di riutilizzo dei dati arriva fino all'80% e ciò garantisce un notevole risparmio.⁶¹ Il web semantico è costituito da un numero crescente di archivi, risultato del lavoro cooperativo di produttori di dati di tutto il mondo; nel momento della creazione dei propri dati, ciascun produttore può utilizzare dati messi in condivisione da altri per identificare l'entità che sta descrivendo e aggiungere i dati di proprio interesse (per esempio, la forma di un nome secondo le regole nazionali, locali o di una particolare comunità).

Vediamo un esempio. Per creare un legame tra una qualsiasi entità (per esempio l'editrice "Tabacco") e un luogo (per esempio "Tavagnacco"), è necessario sapere dove si trova questa località e poterla identificare con certezza. Nel web dei linked data, la ricerca viene effettuata sul web, che rende disponibile un dato di un archivio già esistente, che può fungere da archivio d'autorità condiviso per tutte le agenzie bibliografiche e per qualsiasi altro potenziale utente. Nel caso dell'esempio, si può utilizzare GeoNames, un "archivio d'autorità geografico" nel quale la loca-

⁶⁰ Library Linked Data Incubator Group, *Library Linked Data Incubator Group final report*, paragrafo 2.1.

⁶¹ Stefan Decker et al., *Linked data in the digital humanities*, "Lodlam.net", 2012, p. 19, <<http://lodlam.net/2012/10/27/linked-data-in-the-digital-humanities/>>.

lità è già identificata per tutto il web.⁶² Il lavoro di registrazione dei dati delle località geografiche è già stato fatto e non importa se è stato fatto da catalogatori o a fini bibliografici: i dati esistono già e, purché abbiano le garanzie e le caratteristiche di qualità necessarie, possono e devono essere riutilizzati a fini bibliografici.

Questo approccio non ha un limite; potrebbero essere utilizzati come archivi d'autorità anagrafi di natura completamente diversa: anagrafi cittadine, banche dati delle società proprietarie dei diritti d'autore, sistemi informativi territoriali, anagrafi bancarie, archivi di istituzioni pubbliche, banche dati di beni museali, archivistici, artistici, etc. L'approccio non ha nemmeno una direzione sola, dai dati del web verso le biblioteche. Se le biblioteche pubblicano i propri dati in questa forma, li rendono immediatamente visibili e disponibili ai motori di ricerca e a tutti i potenziali utilizzatori. Un interessante esempio in questo senso è il progetto realizzato al Comune di Firenze, dove i 230 data set creati per lo stradario sono stati incrociati con il VIAF⁶³ e il *Nuovo Soggettario*⁶⁴ e per consentire a un utente di distinguere vie e piazze dedicate a personaggi illustri omonimi.

L'importanza dei linked data per le biblioteche pubbliche e per la pubblica amministrazione e il loro impatto sulla gestione dei dati pubblici sono stati oggetto di un convegno che si è tenuto nel giugno 2012 a Firenze, promosso dall'ICCU e dall'Università di Firenze.⁶⁵

8. Dal catalogo alla navigazione semantica

Al termine di questa ricostruzione dei passaggi che hanno segnato il percorso dei dati dei cataloghi in relazione alla loro capa-

⁶² <<http://www.geonames.org/6542026/tavagnacco.html>>.

⁶³ VIAF: *The Virtual International Authority File*, "Viaf.org", 2012, <<http://viaf.org>>.

⁶⁴ Giovanni Menduni, *Come ordinare gli open data di Firenze con l'aiuto di una biblioteca*, 2012, <<http://www.chefuturo.it/2012/07/come-abbiamo-ordinato-gli-open-data-di-firenze-con-laiuto-di-una-biblioteca/>>.

⁶⁵ *Global Interoperability and Linked Data in Libraries*; special issue, editor: Mauro Guerrini, proceedings by Gianfranco Crupi and Ginevra Peruginelli, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/issue/view/536>>.

cià di rispondere a funzioni di crescente complessità, si aprono diversi interrogativi; in particolare: quali saranno gli effetti dei linked data sulle ricerche degli utenti? Quali su oggetti come la registrazione bibliografica o lo stesso catalogo?

Sul versante utente, il primo effetto dell'esposizione sul web dei dati contenuti nei silos bibliografici dovrebbe essere che la registrazione di un dato oggetto di interrogazione sarà raggiungibile indifferentemente a partire da un'interrogazione sul web con un motore di ricerca o su un OPAC specifico di un servizio bibliotecario. In entrambi i casi, l'utente dovrebbe essere in grado di effettuare con successo – tramite i dati e le relazioni prodotti dalle biblioteche – ricerche consapevoli e ricerche esplorative. La differenza importante tra le due modalità sarà che i linked data creati dalle biblioteche saranno costituiti e presentati secondo standard di qualità e largamente condivisi, offrendo maggiori garanzie di coerenza e autorevolezza in fase di registrazione e di leggibilità in fase di visualizzazione. Dunsire, Hillmann e Phipps infatti affermano che “il vero valore dei linked data bibliografici consiste nella loro coerenza e completezza, importantissimi quando essi sono mescolati nell'ambiente incontrollato del web semantico”.⁶⁶

Il secondo effetto dovrebbe essere che l'utente potrebbe cercare un dato (il nome di un personaggio, il nome di un soggetto etc.) a partire da qualsiasi forma gli sia nota; un utente di qualsiasi parte del mondo e di qualsiasi madrelingua, potrebbe compiere interrogazioni usando termini nella forma *da lui* preferita o conosciuta e trovare *tutte le informazioni* rilevanti (non solo sulle manifestazioni, ma su tutte le entità del Gruppo 1 di FRBR e degli altri Gruppi di entità di FRBR, FRAR e FR SAR). Tale strumento consentirebbe di rispondere alle esigenze della ricerca consapevole: ciò diventa possibile perché tutte le forme attraverso le quali può essere rappresentato il dato ricercato dovrebbero essere collegate tra loro come oggi avviene nel VIAF per le forme autorizzate degli autori nelle diverse regole nazionali.

Un terzo effetto, quando ciò avverrà anche per le forme di un

⁶⁶ Gordon Dunsire, Diane Hillmann, Jon Phipps, *Reconsidering universal bibliographic control in light of the semantic web*, “Journal of Library Metadata”, vol.12, nos. 2/3 (2012), p. 175.

soggetto in un web di dati dove fossero implementate le relazioni semantiche tra concetti (di equivalenza, gerarchiche, associative), sarebbe la navigazione non più tra i documenti del web o le registrazioni bibliografiche, ma tra i concetti-soggetti, sui quali si otterranno contemporaneamente informazioni generali e bibliografiche, rispondendo molto bene ai requisiti della ricerca esplorativa.

Esisteranno ancora la registrazione bibliografica e il catalogo? Nel web dei dati, una registrazione sarà costituita da un insieme di informazioni e dati corredati di un significato non ambiguo relativi a una specifica entità e associati secondo regole *semantiche*, creato a partire dal web dei dati e dotato di collegamenti a partire da un suo qualsiasi componente (e non solo da quelli che oggi costituiscono un accesso ai dati bibliografici secondo ICP). Dal punto di vista della registrazione, l'insieme dei dati necessari è definito in base alle norme che si adottano per l'identificazione (per esempio, gli elementi essenziali in RDA previsti da una agenzia che crea i dati).⁶⁷ I cataloghi che conosciamo oggi avranno il ruolo di luoghi di produzione dei dati di qualità che saranno riversati nel web dei dati per costituirne la parte di maggiore qualità.

Dal punto di vista della visualizzazione o rappresentazione, nel web semantico i dati potranno essere scelti, presentati e disposti in modi diversi, e dipendenti dal *contesto* di visualizzazione: nel web semantico il contesto è sovrano.⁶⁸ Per esempio, la forma con la quale comparirà il nome di un autore russo, dipenderà dal contesto nel quale dev'essere rappresentato il dato (ovvero il contesto per un utente russo, inglese, italiano, francese, etc.). Oppure, il numero, l'ordine e la forma degli elementi per la rappresentazione di una manifestazione dipenderanno dal contesto: i dati disponibili nel web verranno selezionati in modo differente attraverso un *application profile* per il sito di una libreria online, per i riferimenti bibliografici di una voce di *Wikipedia*, per la descrizione di una risorsa in un catalogo in linea, etc. La rappresentazione di una qualsiasi entità potrà assumere la forma di una pagina web

⁶⁷ Vedi, per esempio, Library of Congress, *LC RDA core elements*, 2014, <http://www.loc.gov/aba/rda/pdf/core_elements.pdf>.

⁶⁸ Kingsley Idehen, *Creating, Deploying and Exploiting Linked Data*, 2012, p. 35, <[http://virtuoso.openlinksw.com/presentations/Creating_Deploying_Exploiting_Linked_Data2/Creating_Deploying_Exploiting_Linked_Data2_TimBL_v3.html#\(1\)>](http://virtuoso.openlinksw.com/presentations/Creating_Deploying_Exploiting_Linked_Data2/Creating_Deploying_Exploiting_Linked_Data2_TimBL_v3.html#(1)>).

che visualizza molti dati su quella entità e ne evidenzia le sue relazioni. La prima parte costituisce, quindi, la descrizione, costruita *on the fly*, dell'entità; la seconda parte costituisce lo strumento per la navigazione a partire da quella entità, per garantire all'utente la possibilità di effettuare ricerche esplorative e la scoperta casuale di connessioni (serendipità).

Due esempi della forma che potranno assumere le presentazioni dei dati si trovano in *data.bnf.fr* e *datos.bne.es*; nelle figure 6 e 7 si vedono le rispettive pagine web per Wolfgang Amadeus Mozart:



Figura 6 – Wolfgang Amadeus Mozart su data.bnf.fr (http://data.bnf.fr/14027233/wolfgang_amadeus_mozart/)



Figura 7 – Wolfgang Amadeus Mozart su datos.bne.es (<http://datos.bne.es/autor/XX1147441.html>)

Dal punto di vista della registrazione fisica dei dati, secondo Martin Malmsten la registrazione bibliografica si disintegra nel web semantico,⁶⁹ perché una parte sempre più consistente dei dati che costituiranno le registrazioni saranno estranei al catalogo prodotto *in house* da una biblioteca o da un sistema bibliotecario. Quindi nella logica della distribuzione del lavoro di costruzione dei linked data, parte dei dati *d'interesse bibliografico* non avranno più *origine bibliografica*, ma potranno essere prodotti altrove: da agenzie del ciclo del libro, come editori, librai, etc. e da agenzie con obiettivi completamente diversi, come per esempio i dati geografici o i dati di identificazione degli organi amministrativi di uno stato. Come ha ricordato Karen Coyle al convegno sui linked data di Firenze, sarà indispensabile che le biblioteche superino la logica del possesso anche per i dati, e non più solo per le raccolte.⁷⁰

Si apre il problema della *qualità* dei dati; è un aspetto che ha molte sfaccettature. Il problema si pone dal punto di vista teorico, in relazione alla definizione del concetto di qualità dei dati⁷¹ e si pone dal punto di vista pratico: l'esperienza della Bibliografia nazionale svedese indica addirittura che la strada sarà il trattamento delle proprie risorse interne come se fossero esterne; il lavoro di authority potrebbe trasformarsi dal controllo locale dei dati alla verifica della qualità e della coerenza di dati e archivi esterni.⁷² In questo nuovo contesto le biblioteche avranno il compito di individuare, tra le molte fonti di dati disponibili, quelle più autorevoli e collegare dati tra loro coerenti: è una competenza che i bibliotecari hanno da sempre.

Si pone, infine, da un punto di vista culturale ed etico: “Chi

⁶⁹ Martin Malmsten, *Cataloguing in the open. The disintegration and distribution of the record*, “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013), p. 417-423, <<http://leo.cilea.it/index.php/jlis/article/view/5512>>.

⁷⁰ Karen Coyle, *Linked data: an evolution*, “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013), p. 53-61, <<http://leo.cineca.it/index.php/jlis/article/view/5443>>.

⁷¹ Seth Van Hooland, Ruben Verborgh, *Linked data for libraries, archives and museums: how to clean, link and publish your metadata*, London: Facet Publishing, 2014, p. 73-75.

⁷² Martin Malmsten, *Cataloguing in the Open. The disintegration and distribution of the Record*.

gestisce tecnicamente, semanticamente e linguisticamente vocabolari e ontologie svolge un ruolo determinante nella definizione dei lemmi e delle relazioni tra di essi; le voci, infatti, saranno utilizzate automaticamente e, dunque, acriticamente dai processi inferenziali compiuti dalle macchine. Le scelte nelle definizioni di nuovi vocabolari e ontologie o nelle loro estensioni assumono, pertanto, una dimensione tecnica e culturale rilevante nel processo della comunicazione globale”.⁷³ Su questo problema la comunità professionale bibliotecaria dovrà assumere un ruolo fondamentale; non solo per la sua lunga tradizione di valori fondamentali come la libertà intellettuale, la privacy, l’equità nell’accesso alla conoscenza e all’informazione, il diritto a ricercare l’informazione che si preferisce e la democrazia, ma soprattutto perché è la sola che possiede oggi le competenze necessarie per garantire la *qualità* dei dati.

⁷³ Carlo Bianchini, Mauro Guerrini, *Introduzione a RDA*, cit., p. 19.

APPENDICE

VOCABOLARIO O SCHEMA LOD-READY PER L'INTERLINKING CON METADAT RESPONSIBLE BODY E SUBJECT E LA LORO ABILITAZIONE COME LOD

Nelle tabelle che seguono¹ vengono presentati alcuni vocabolari controllati e schemi semantici (classificazioni, ontologie) le cui entità costitutive sono identificate con gli indirizzi web, quindi con i valori non literal in riferimento ai corrispondenti valori *literal* (stringhe di soggetto, classi, voci di thesauri). L'implementazione di tali KOS in un repository rappresenta un approccio innovativo per il controllo d'autorità e per l'arricchimento semantico delle proprietà di alcuni metadati (negli esempi in tabella per *Responsible Body* e *Subject*); favorisce, inoltre, la creazione degli *accessi multifocali*, ovvero degli accessi unici proiettati a una diversità di risorse distribuite in modo trasversale sul web che hanno le unità informative (dati di contenuto) in comune, in corrispondenza alla logica della modellazione descritta nelle *Raccomandazioni LODE-BD* e alla prospettiva della creazione di un *International LOD library ecosystem*² e di un globale *Open repository ecosystem*.³

¹ Iryna Solodovnik è responsabile della stesura dell'Appendice.

² Vedi: SWIB 12. *Semantic web in libraries. Towards an international LOD library ecosystem*. 26 - 28 November 2012, Cologne, Germany, <<http://swib.org/swib12/>>, <Slides and videos online>.

³ Vedi: *Towards repository ecosystems. open repositories 2014. The Ninth International Conference on Open Repositories, OR2014, will be held 9-13 June 2014 in Helsinki, Finland*, <<http://aims.fao.org/community/metadata/blogs/open-repositories-2014-call-proposals-towards-repository-ecosystems>>.

Schemi a supporto della codifica del metadato RESPONSIBLE BODY	
Virtual International Authority (VIAF), < http://viaf.org/ >	Progetto congiunto di diverse biblioteche nazionali e alcune agenzie di biblioteche regionali e transnazionali, sostenuto dall'OCLC. VIAF ha l'obiettivo di ridurre i costi e aumentare l'utilità degli authority file per le piattaforme digitali, tramite i corrispondenti collegamenti agli <i>Authority name files</i> per l'uso normalizzato dei nomi degli autori identificati con i valori non literal URIs e pubblicati nel web come entità linked open data.
MARC code list for relators, < http://id.loc.gov/vocabulary/relators.html >	Lista di termini che descrivono i ruoli degli autori, con i codici loro associati per creare le relazioni tra i nomi e le risorse.
ORCID (Open Researcher and Contributor ID), < http://orcid.org/ >	Registro avviato nel 2010 come iniziativa no-profit; consente la registrazione degli identificatori univoci (alfanumerici) dei ricercatori. ORCID risponde al problema di collegare in modo attendibile gli autori alle opere. Esso è più di un registro: esprime la collaborazione di una comunità di ricercatori per integrare gli identificatori nel workflow della ricerca. La registrazione degli ID è gratuita per gli utenti singoli.
FOAF (Friend Of A Friend), < http://xmlns.com/foaf/0.1/ >	Ontologia per descrivere persone, con le loro attività e le relazioni con altre persone e oggetti, utilizzabile da chiunque. FOAF permette a gruppi di persone di partecipare ai social network senza la necessità di accedere a un database centralizzato. FOAF è un vocabolario descrittivo espresso tramite il modello di dati RDF e OWL. I computer possono utilizzare FOAF, per esempio, per cercare tutte le persone che vivono in Europa, o tutte le persone che hanno un amico in comune; ciò perché permette di definire le relazioni tra persone. Ogni profilo ha un identificativo univoco che viene utilizzato quando vengono definite le relazioni; per esempio, l'indirizzo email, l'URI dell'homepage o del blog della persona.

Schemi a supporto della codifica del metadato Subject	
Dewey Decimal Classification (DDC), < http://dewey.info/ >	OCLC ha rilasciato il set completo di DDC23 come linked open data tramite il servizio Dewey.info, uno spazio online di sperimentazione per DDC in linked data: oltre 23.000 numeri assegnabili e didascalie in inglese.
Library of Congress (LC) Classification (LCC), < http://id.loc.gov/ >	Sistema di classificazione sviluppato dalla Library of Congress disponibile come linked data tramite il servizio LC's Linked Data Service (ID.LOC.GOV).
Library of Congress (LC) Subject Headings (LCSH), < http://lcsjsh.org/ >	Versione SKOS di LCSH disponibile come linked data tramite il servizio LC's Linked Data Service (ID.LOC.GOV); è costituita da 30.000 voci monolingue (in inglese).
Unesco Thesaurus, < http://skos.um.es/unescothes >	La versione in SKOS è disponibile in quattro lingue: inglese, spagnolo, francese e russo. Unesco Thesaurus è strutturato in sette settori suddivisi in micro-thesauri, che riguardano: istruzione, cultura, scienze naturali, scienze umane e sociali, comunicazione e informazione.
Thesaurus del <i>Nuovo Soggettario</i> in formato SKOS/RDF, < http://thes.bncf.firenze.sbn.it/thes-dati.htm >	Il <i>Nuovo Soggettario</i> , in quanto linguaggio documentario, si configura come un sistema integrato di quattro componenti progettate per poter operare insieme al fine di supportare il processo d'indicizzazione: un thesaurus multidisciplinare in lingua italiana corredato da equivalenti inglesi previsti da Library of Congress Subject Headings (LCSH); un apparato normativo inerente controllo terminologico e costruzione delle stringhe di soggetto; un corredo applicativo costituito da un Manuale e da note sintattiche associate ad alcuni termini; gli archivi delle stringhe di soggetto prodotte da chi impiega il linguaggio. Il <i>Nuovo Soggettario</i> si pone nella prospettiva di sviluppare sempre più il versante dell'interoperabilità con altri strumenti di organizzazione della conoscenza. Per questo, i metadati del thesaurus sono

	disponibili, oltre che tramite il protocollo Zthes, anche nel formato SKOS/RDF nell'ottica del web semantico. La Biblioteca nazionale centrale di Firenze ha in corso collaborazioni a livello nazionale e internazionale per continuare ad ottimizzare la pubblicazione, come linked data, dei metadati prodotti.⁴
EuroVoc Thesaurus, <http://datahub.io/dataset/eurovoc-in-skos>	Thesaurus multidisciplinare e multilingue, disponibile nelle 23 lingue ufficiali dell'Unione europea; comprende la terminologia dei settori d'attività dell'Unione europea, con un'attenzione particolare per l'attività parlamentare. Il thesaurus è basato sulle ontologie e sulle tecnologie del web semantico, in linea con le raccomandazioni del W3C e con gli ultimi sviluppi degli standard di classificazione.
AGROVOC Thesaurus, <http://datahub.io/dataset/agrovoc-skos>	Thesaurus multilingue (22 lingue); riguarda i campi tematici dell'agricoltura, selvicoltura, pesca, sicurezza alimentare e di settori affini (con oltre 30.000 voci). AGROVOC incoraggia gli sviluppatori dei sistemi di gestione dell'informazione agricola a incorporare il thesaurus nelle loro applicazioni, anziché utilizzare thesauri locali.
Library of Congress' Thesaurus for Graphic Materials: <http://id.loc.gov/vocabulary/graphicMaterials.html>	Strumento impiegato per l'indicizzazione delle risorse visive: fotografie, stampe, disegni di progettazione etc.; contiene oltre 7.000 voci. Aggiornato costantemente.
STW (Standard-Thesaurus Wirtschaft): Thesaurus for Economics, <http://datahub.io/dataset/stw-thesaurus-for-economics> <http://zbw.eu/beta/sparql/>	Thesaurus bilingue (inglese, tedesco) che fornisce il vocabolario nel campo dell'economia: circa 6.500 voci standardizzate e circa 18.000 termini per definire l'accesso semantico e normalizzato alle risorse.

⁴ La voce è redatta da Elisabetta Viti.

GEMET (the GEneral Multilingual Environmental Thesaurus), < http://taskman.eionet.europa.eu/projects/zope/wiki/GEMETLinkedData >	Copre il dominio dell'ambiente; disponibile in 29 lingue. La versione SKOS è costituita da oltre 5.000 voci.
Getty Thesaurus of Geographic Names (TGN), < http://datahub.io/dataset/getty-tgn >	Vocabolario strutturato di nomi geografici e di altre informazioni per l'indicizzazione delle risorse nei settori arte e architettura.
Alcune ontologie <i>LOD-ready</i>	
VIVO Ontology, < http://datahub.io/dataset/vivo >	L'ontologia VIVO è stata concepita per creare un facebook semantico per gli scienziati tramite un'ontologia di base: modellazione dei nomi dei ricercatori, dei titoli delle pubblicazioni, delle località dei servizi, etc. La versione 1.0 contiene 236 classi con 278 proprietà e 222 proprietà dei dati; incorpora le classi da ontologie come BIBO, Dublin Core, FOAF e SKOS. I dati VIVO sono focalizzati sulla ricerca (titoli di studio, pubblicazioni, esperienza, borse di studio), sulla didattica (seminari, corsi di formazione), sui servizi (organizzazione di conferenze e altri servizi).
Bibliographic Ontology (BIBO), < http://bibliontology.com/specification >; < http://purl.org/ontology/bibo/ >	Fornisce i concetti e le proprietà principali per la descrizione di citazioni e riferimenti bibliografici (citazioni, libri, articoli) nel web semantico. Altri vocabolari (elementi di metadati DC per una descrizione bibliografica, FOAF) possono essere combinati con i termini BIBO, come estensioni locali.
GeoNames Semantic Web, < http://datahub.io/it/dataset/geonames-semantic-web >	GeoNames permette di aggiungere informazioni semantiche geospaziali alle risorse web; contiene oltre 6.2 milioni di toponimi che hanno un identificatore di localizzazione URL univoco, con un servizio web RDF corrispondente. Descrive le relazione tra toponimi.

SURF Classification Schemes https://wiki.surfnet.nl/display/standards/info-eu-repo#info-eu-repo-ClassificationSchemes	
<i>Namespace/ dominio del vocabolario nel web</i>	Vocabolario come schema di codifica
<info:eu-repo/classification/ddc>	Dewey Decimal Classification (DDC): < http://www.oclc.org/dewey/ >
<info:eu-repo/classification/udc>	Universal Decimal Classification: < http://www.udcc.org/http://www.udcc.org/ >
<info:eu-repo/classification/lcc>	Library of Congress Classification: < http://www.loc.gov/catdir/cpsol/lcco/http://www.loc.gov/catdir/cpsol/lcco/ >
<info:eu-repo/classification/lcsh>	Library of Congress Subject Headings (LCSH)
<info:eu-repo/classification/swd>	Schlagwortnormdatei (SWD)
<info:eu-repo/classification/bicssc>	Book Industry Communication - Standard Subject Categories: < http://www.bic.org.uk/7/BIC-Standard-Subject-Categories/ >
<info:eu-repo/classification/nbc>	Nederlandse Basisclassificatie (Dutch Basic Classification): < http://www.kb.nl/vak/basis/bc04.pdf >
<info:eu-repo/classification/jel>	Sistema di classificazione "Journal of Economic Literature" (JEL). Gli articoli su rivista di scienze economiche vengono ampiamente classificati secondo i codici della Classificazione JEL: < http://www.aeaweb.org/journal/jel_class_system.php >

Esperienze di digital libraries che hanno accolto la pratica operativa di LOD

Tra le varie esperienze di digital libraries che hanno accolto la pratica operativa di LOD⁵ si segnala Europeaana, la cui infrastruttura gestionale (impostata secondo il modello dati Europeaana Data Model) è potenziata dai meccanismi che arrecano il valore aggiunto alle risorse disseminate tramite i collegamenti creati con le risorse d'autorità LOD, come Geonames, VIAF, GND (*Gemeinsame Normdatei - Integrated Authority File of the German National Library*),⁶ Geonames, LCSH.

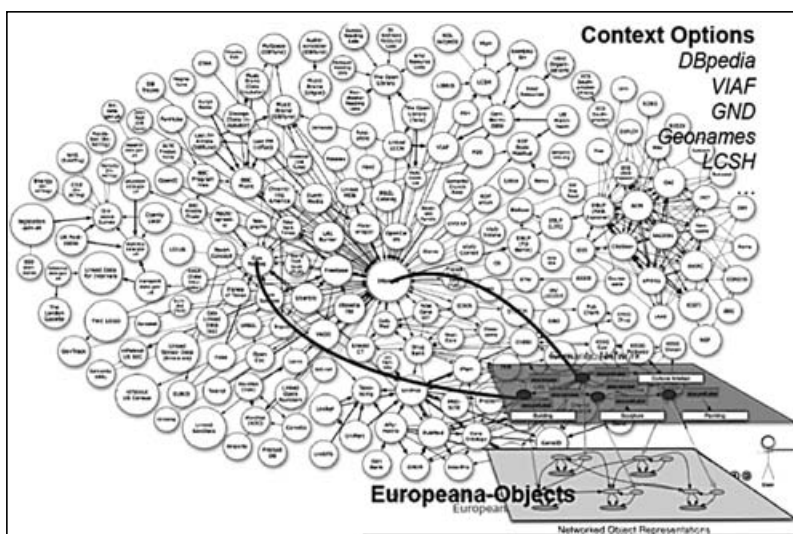


Figura 1 – Collegamenti tra Europeana objects e dataset d'autorità disponibili come LOD⁷

⁵ Carl Jay Lagoze, *Lost identity: the assimilation of digital libraries into the web*. PhD thesis, Information Science, Cornell University, 2010, <<http://carllagoze.files.wordpress.com/2012/06/carllagoze.pdf>>.

⁶ <<http://d-nb.info/standards/elementset/gnd>>.

⁷ *The Europeaana Data Model: constraints and opportunities*. Europeaana Connect, Stefan Gradmann, based on work done with Martin Dörr, Hennie Steffen, Antoine Isaac, Carlo Meghini, Guus Schreiber, Herbert Van de Sompel, slide 6, <http://www.europeanaconnect.eu/europeanatech/documents/Gradmann_Stefan_LinkedOpenEuropeana.pdf>.

Un altro esempio di come può essere creato valore aggiunto tramite i collegamenti con dataset LOD è il servizio di ricerca di contenuti digitali offerto dalla piattaforma OpenAGRIS,⁸ la quale permette di cercare e di recuperare le risorse informative ospitate entro la medesima piattaforma e di rilevare le risorse correlate alla tematica da piattaforme esterne. Questo tipo di accesso multifocale ai contenuti correlati è possibile grazie alle relazioni create tra valori di proprietà di metadati che definiscono il tema (soggetto) di risorse informative con voci dell'AGROVOC Thesaurus. Le entità dell'AGROVOC Thesaurus, a loro volta, sono collegate con le entità di altri vocabolari controllati, anche essi disponibili sul web come LOD.

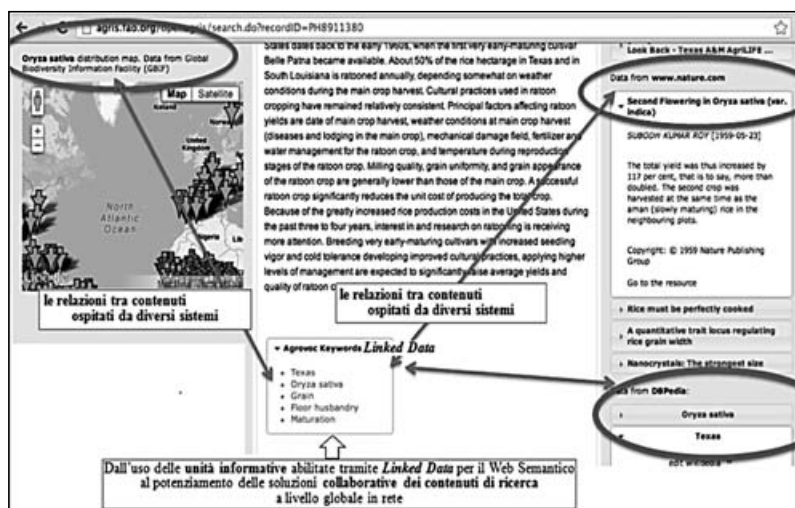


Figura 2 - Impiego del Thesaurus LOD AGROVOC nella piattaforma di ricerca OPENAGRIS

⁸ OpenAGRIS (applicazione web per aggregare le risorse da diverse fonti Web con più di 60 milioni di triple RDF): <<http://aims.fao.org/openagris>>; <<http://agris.fao.org/agris-search/search/display.do?f=2010%2FJFP%2FJFP1014.xml%3BJP2010001379>>; vedi, inoltre: Stefano Anibaldi, Yves Jaques, Fabrizio Celli, Armando Stellato, Johannes Keizer, *Migrating bibliographic datasets to the Semantic Web: the AGRIS case*, "Semantic web journal" 00 (20xx), p. 1-8: <<http://iospress.metapress.com/content/115562xk70234n79/fulltext.pdf>>; Fabrizio Celli, Stefano Anibaldi, Maria Folch, Yves Jaques, Johannes Keizer, *OpenA-*

Esistono tre opzioni per l'esposizione di contenuti informativi sul web semantico:

- 1) *pushing*: la pubblicazione di contenuti informativi più rappresentativi come LOD (metadati, vocabolari controllati, authority file, altre risorse e loro parti costitutive), con la loro successiva registrazione tramite DataHub CKAN e il collegamento con altri dataset presenti nella LOD cloud;
- 2) *pulling*: l'affiancamento di contenuti più rappresentativi da dataset disponibili tramite CKAN e presentati in LOD Cloud: DBpedia, Eurovoc Thesaurus, UNESCO Thesaurus, AGROVOC Thesaurus, Classificazione Dewey, LCSH, RAMEAU Subject Headings;⁹
- 3) combinazione delle opzioni (1) e (2).

GRIS: using bibliographical data for linking into the agricultural knowledge web, risorsa prodotta per *FAO of the United Nations*, 2011: <<http://www.fao.org/docrep/015/an642e/an642e00.pdf>>.

⁹Library Linked Data Incubator Group, *Library Linked Data Incubator Group final report*: <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>; Fabrizio Celli, Yves Jaques, Stefano Anibaldi, Johannes Keizer, Pushing, Pulling, Harvesting, Linking: Rethinking Bibliographic Workflows for the Semantic Web, EFITA-WCCA-CIGR Conference "Sustainable Agriculture through ICT Innovation", Turin, Italy, 24-27 June 2013, <http://eprints.rclis.org/21112/1/2013_EFITA%20Pushing_Pulling.pdf>.

GLOSSARIO ESSENZIALE

Per alcune definizioni vedi anche il paragrafo 4.1.1.

API (Application Programming Interface). Insieme di funzioni applicative per sviluppare programmi che devono dialogare con il sistema operativo o con altre applicazioni distribuite in rete. Lo sviluppo delle web-API ha comportato un'esplosione di applicazioni specializzate che combinano dati provenienti da fonti diverse. Sono di solito invisibili all'utente; consentono l'interazione di applicazioni web o mobile, con un determinato servizio o con repository di dati.¹ La necessità di usare API specializzate per i differenti servizi e set di dati (per esempio, i vari tipi di Google API, le Facebook API, le WorldCat Search API, etc.) richiedono un notevole sforzo ai programmatori che devono conoscere le funzioni disponibili per utilizzare in modo appropriato ciascuna API e devono scrivere un codice personalizzato per l'accesso ai dati di ciascuna fonte. I web services sono un approccio più evoluto e flessibile che, avvalendosi degli standard del web, permettono di sviluppare applicazioni distribuite ed interoperanti secondo una logica orientata ai servizi.

Blank node. Raggruppamento di dati utilizzato in forma di soggetto o di oggetto per esprimere elementi composti di una risorsa; per esempio, l'*indirizzo* utilizzato come blank node, che rimanda ad attributi diversi quali la via, il numero civico, la città.²

CMS (Content Management System) o WCM (Web Content Management). Sistema di gestione di contenuti, ovvero di metodi e tecniche per automatizzare i processi di creazione, gestione e distribuzione di contenuti attraverso il web. Molti portali sono basati su CMS open source, per la realizzazione delle funzioni di gestione di contenuti. Questi sistemi sono utili qualora i fruitori

¹ <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

² <<http://www.w3.org/TR/REC-rdf-syntax/>>.

dei dati siano solamente utenti umani. I tre prodotti principali e più utilizzati sono Wordpress, Drupal e Joomla, sufficientemente flessibili e personalizzabili da adeguarsi alla pubblicazione di (linked) open data; in particolare Drupal include in maniera nativa le tecnologie legate al web semantico, come RDF e RDFa.³

Crawler. Detto anche web crawler, spider o robot, è un software che analizza i contenuti esposti da siti web e portali (o di database) in un modo automatizzato, in genere per conto di un motore di ricerca.⁴

Creative Commons Public Licenses (CCPL). Sono licenze d'uso nate per ampliare le utilizzazioni consentite nell'ambito della legislazione in materia di diritto d'autore. Le CCPL, infatti, rendono semplice, per il titolare dei diritti d'autore, segnalare in maniera chiara che la riproduzione, diffusione e circolazione della propria opera è esplicitamente permessa.⁵ È possibile utilizzare una versione RDF del vocabolario, soprattutto quando si gestiscono linked data.⁶

Crypto, Digital Signature e Security. Standard e tecnologie di crittografia e di firma digitale che permettono di aumentare il livello di sicurezza dei linked data scambiati e di verificare l'integrità e la provenienza delle asserzioni del web semantico.

CKAN Data Hub. Data management system open source per la pubblicazione, la condivisione, la ricerca e la fruizione di dati, in modalità aperta.⁷ Il sistema CKAN, sviluppato dall'Open Knowledge Foundation, è stato concepito per la descrizione di risorse recuperate come file accessibili tramite URL. È una piattaforma onnicomprensiva, ben integrata e altamente perso-

³ <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

⁴ <<http://it.wikipedia.org/wiki/Crawler>>; <<http://www.treccani.it/vocabolario/crawler/>>.

⁵ <<http://www.creativecommons.it/cosa-fa-cc>>.

⁶ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio-pubblicolg2014_v0.7finale.pdf>.

⁷ <<http://datahub.io/>>.

nalizzabile, con cui si possono realizzare tutti gli elementi di un sistema di gestione di open data: memorizzazione fisica, organizzazione logica, creazione di metadati ed esposizione su un sito web.⁸

Data interchange/RDF. Data model utilizzato per la strutturazione di linked data.

Data mining. Processo di estrazione di elementi di tipo cognitivo da banche dati di grandi dimensioni tramite l'applicazione di algoritmi che individuano le associazioni nascoste tra le informazioni e le rendono visibili. Gli algoritmi di data mining sono stati sviluppati per far fronte all'esigenza di sfruttare il patrimonio informativo contenuto nelle grandi raccolte di dati, per il suo utilizzo scientifico, industriale od operativo.⁹ I processi di data mining elaborano grandi quantità di dati, provenienti, da fonti diverse, che rappresentano i "fatti" (numeri, metadati, report statistici, etc.) del contesto esaminato (per esempio, la produzione e il commercio di alcune tipologie di prodotti in un determinato ambito territoriale), dai quali estraggono informazioni (per esempio, il comportamento e le preferenze degli acquirenti) che poi sono trasformate in "conoscenze" sui comportamenti passati e sulle tendenze (per esempio, per elaborare previsioni sulle offerte da proporre).

Dataset. Collezione di dati pubblicati e mantenuti da un singolo fornitore, disponibile come RDF e consultabile, per esempio, tramite deferenziali HTTP URI o uno SPARQL endpoint. È un blocco di dati provenienti da sorgenti esistenti con una loro struttura o aggregazione affine (un'affinità disciplinare, cronologica, tecnologica) che deriva dalla sorgente di provenienza. I dati vengono erogati e gestiti congiuntamente. Un dataset elenca i valori per ciascun attributo, per esempio, il nome e il cognome di una persona, il colore e il peso di un oggetto; ogni valore è definito *dato*. Un dataset può assumere la forma di sequenza di

⁸ <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

⁹ <<http://www.cineca.it/it/content/data-mining>>.

caratteri marcati mediante tag, come, per esempio, un file XML. In analogia con questo concetto si parla di *dataset RDF*, ovvero di una collezione di risorse (o di grafi) accomunate da una caratteristica e identificate da URI dereferenzabili.

Dati grezzi (raw data). Dati raccolti che non hanno subito alcuna modifica, manipolazione o aggregazione.¹⁰

Dato. Rappresentazione fisica di fatti, eventi atomici, fenomeni oggettivi, informazioni atte alla comunicazione, interpretazione ed elaborazione da parte di essere umani o mezzi automatici.¹¹ Un *dato* (dal latino *datum*, che significa letteralmente fatto) è una descrizione elementare, codificata, di una cosa, una transazione, un avvenimento. L'elaborazione dei dati può portare a ricavare informazioni. I dati possono presentarsi sotto diverse forme: numeri e lettere dell'alfabeto (testo), immagini statiche, immagini in movimento (video), suono (audio) e altro; possono essere conservati su diversi mezzi (o supporti) fisici (carta, disco magnetico o ottico, come CD, DVD, chiavette USB, etc.) e trasmessi tramite una rete di telecomunicazioni tra più utenti. Nei sistemi informatizzati l'elaborazione tramite programmi viene effettuata sulla rappresentazione binaria (sequenze di 0 e 1) dei dati, mentre i caratteri dei dati testuali (composti da sequenze alfanumeriche) sono codificati in base a standard internazionali come Unicode e ASCII.

Dump. Il dump viene utilizzato come salvataggio logico del database che, in caso di anomalie dei sistemi o dei supporti fisici, può essere utilizzato per ripristinare il database originale mediante corrispondenti procedure di ripristino (operazione di *restore*). Nei data base relazionali è un file che contiene un riepilogo della struttura delle tabelle del database medesimo e dei relativi dati ed è normalmente nella forma di una lista di dichiarazioni SQL.

¹⁰ <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

¹¹ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicolg2014_v0.7finale.pdf>.

Dump RDF. File prodotto da un dataset RDF, messo a disposizione nel web per il download e utilizzabile per il trasferimento o il ripristino di interi dataset.

Dereferenziare. L'azione di recuperare la rappresentazione di una risorsa identificata da un URI (dereferencing the URI). Le pagine e i documenti familiari per gli utenti del web sono risorse informative, identificate da un URI. Le risorse hanno tipicamente una o più rappresentazioni che possono essere recuperate tramite il protocollo HTTP. Il messaggio di ritorno di una richiesta HTTP veicola questa rappresentazione. Per esempio, cercando in WorldCat il titolo *Il nome della rosa* si ottiene una rappresentazione breve e una più estesa in formato HTML della corrispondente risorsa bibliografica, ma si possono, inoltre, ottenere la rappresentazione nel formato citazionale EndNote e in RDF sia in un formato tabellare leggibile sul browser, sia in uno dei vari tipi di rappresentazione *content-negotiable* previsti per RDF: Turtle, JSON-LD, RDF/XML, N-TRIPLES, HTML+RDFa.

FOAF (Friend of a Friend). Ontologia utilizzata per descrivere le persone, con le loro attività e le relazioni con altre persone e oggetti, e in particolare predicati quali *foaf:knows*, *foaf:based_near*, *foaf:topic_interest*.¹²

GRDDL (Gleaning Resource Description from Dialects of Language). Meccanismo raccomandato dal W3C; consente di estrarre RDF da pagine web e, quindi, di ottenere triple RDF da documenti XML e XHTML. Il risultato in RDF può essere memorizzato in un database e acceduto tramite SPARQL.

HTML (Hyper Text Markup Language). Linguaggio usato per la formattazione di documenti ipertestuali disponibili nel WWW sotto forma di pagine web. È un linguaggio di pubblico dominio, la cui sintassi è stabilita dal W3C. HTML descrive le modalità di visualizzazione grafica (layout) del contenuto di una pagina web tramite *tag di formattazione*; supporta l'inserimento

¹² <www.foaf-project.org>.

di script (porzione di codice che viene eseguita dal browser) e oggetti esterni, quali immagini, filmati e sequenze audio. Il linguaggio HTML ha lo scopo di gestire i contenuti associandone o specificandone allo stesso tempo il layout all'interno della pagina web da realizzare grazie all'utilizzo di *tag* diversi.

Inferenza. Processo logico tramite cui un'informazione viene dedotta da altre informazioni,¹³ che permette cioè, a partire da una proposizione accolta come vera, di passare a una seconda proposizione, la cui verità è dedotta dal contenuto della prima.

Per esempio, date le due asserzioni seguenti:

- a) tutti i filosofi sono persone
- b) Immanuel Kant è un filosofo

Il processo inferenziale consiste nel dedurre, pur in assenza di altre informazioni, che

- c) Immanuel Kant è una persona.

La terza asserzione non è, dunque, presente nei dati, bensì è dedotta dalle altre due asserzioni.

Interlinking dei dati. Link tra risorse appartenenti a dataset esterni. Il linking è una funzionalità molto importante per i linked data e può essere considerata una forma particolare di arricchimento. L'arricchimento avviene grazie all'interlinking tra dataset di origine diversa, tipicamente tra istituzioni diverse, ma può avvenire anche all'interno della stessa istituzione.¹⁴

Interoperabilità. Capacità di sistemi differenti e autonomi di cooperare e di scambiare informazioni in maniera automatica, sulla base di regole condivise.¹⁵ L'interoperabilità si basa sulla standardizzazione dei protocolli di comunicazione, dei formati dei dati e dei servizi che consentono alle applicazioni di interagire tra loro per realizzare determinati compiti; ciò è

¹³ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

¹⁴ <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

¹⁵ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

una caratteristica indispensabile dei sistemi cooperativi distribuiti in rete.

Interoperabilità semantica. Capacità di elaborare informazioni da fonti esterne o secondarie senza perdere il significato reale delle informazioni nel processo di elaborazione.¹⁶

JSON-LD (JavaScript Object Notation - Linked Data). JSON è un formato aperto e standardizzato dall'Internet Engineering Task Force (IETF), come RFC 7159. Usa del testo umanamente leggibile per trasportare i dati come coppie attributo-valore. JSON-LD (LD sta per Linked Data) usa JSON per trasportare i linked data.¹⁷ Viene proposto dal W3C come formato di serializzazione per i linked data. Il gruppo di lavoro che l'ha definito ha posto due obiettivi: 1) mettere a disposizione un'ulteriore funzionalità al framework RDF; e, 2) avvicinare il mondo dello sviluppo web e degli utilizzatori dei sistemi di gestione dati NoSQL. L'iniziativa intende semplificare l'integrazione con i sistemi esistenti che già usano la sintassi JSON in modo da promuovere l'utilizzo dei linked data negli ambienti di programmazione di tipo web, favorire lo sviluppo di servizi web interoperabili e la memorizzazione dei linked data nei sistemi di archiviazione basati su JSON per ridurre il gap esistente per l'adozione di RDF nei tradizionali contesti web (per esempio, la proposta *Schema.org*).¹⁸

Library linked data (LLD). Dataset sviluppati, pubblicati e condivisi per supportare l'apertura e l'interoperabilità dei dati nel mondo delle biblioteche. Essi sono basati sugli standard, sulle esperienze e sulle norme di buona pratica per l'adozione di tecnologie del web semantico in questo settore. Il paradigma è finalizzato all'arricchimento e all'interconnessione dei dati; offre soluzioni innovative a molti problemi attuali delle biblioteche,

¹⁶ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

¹⁷ Definito in *JSON-LD 1.0. A JSON-based Serialization for Linked Data*, <<http://www.w3.org/TR/json-ld/>>.

¹⁸ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

tra i quali la ricerca avanzata nel web, il controllo d'autorità, la classificazione, la portabilità dei dati e la disambiguazione.¹⁹

Licenza. Il contratto, o altro strumento negoziale, con il quale si regolano i diritti relativi all'uso del software e dei dati, esplicitandone le condizioni d'uso e di distribuzione.²⁰

Linked data. Modalità di pubblicazione di dati strutturati e connessi, collegati tra loro. Si basa su tecnologie e standard web aperti come HTTP e URI e ne estende l'applicazione per fornire informazioni che possano essere lette e comprese da macchine. Ciò rende possibile collegare e utilizzare dati provenienti da diverse sorgenti. Secondo il gruppo di ricerca della Freie Universität di Berlino, il *web di dati* è costruito su due semplici idee:

- 1) utilizzare RDF per pubblicare dati strutturati nel web;
- 2) utilizzare HTTP URI per costruire link RDF tra dati appartenenti a basi dati differenti.

In sintesi: pubblicare dati in RDF e collegarli tra loro.²¹

Linked open data (LOD). Sono linked data aperti, o, più esattamente, linked data liberamente utilizzabili grazie a licenze open e rappresentati secondo sintassi aperte.²²

Linkset. Collezioni di triple RDF in cui il soggetto e l'oggetto della tripla sono descritti in differenti dataset. Un linkset in voID è una sottoclasse di un dataset, utilizzato per memorizzare triple che esprimono la relazione del link tra due dataset. In ogni tripla di link il soggetto è una risorsa che risiede in un dataset e l'oggetto è una risorsa che risiede in un altro dataset. Questo modello permette un modo flessibile e potente di stabilire un

¹⁹ <<http://www.linkedheritage.eu/linkeddataseminar/>>.

²⁰ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_publicolg2014_v0.7finale.pdf>.

²¹ <<https://atcult.wordpress.com/2011/01/14/biblioteche-e-linked-data/>>.

²² <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>. Vedi, inoltre: Mirna Willer, Gordon Dunsire, *Bibliographic information organization in the semantic web*, cit., p. 132.

dialogo molto dettagliato sul link tra due dataset, per esempio su quanti link o quali tipi di link sono presenti (per esempio, *owl:sameAs* o *foaf:knows*), o chi reclama queste dichiarazioni.

Literal. Simbolo che rappresenta se stesso, una notazione usata per rappresentare una costante, il cui valore può essere un numero, un carattere, una sequenza di caratteri (stringa) o un booleano (0 o 1). In RDF i literal sono usati per rappresentare valori come stringhe, numeri e date. Un literal può essere l'oggetto di una tripla e può essere usato per descrivere le proprietà delle risorse, per esempio, il nome di una persona, il titolo di un'opera, la data di nascita di un autore.

Mashup. Applicazione che usa contenuto da più sorgenti per creare un servizio completamente nuovo. Il contenuto dei mashup è normalmente preso da terzi via API, tramite feed (per esempio, RSS e Atom) o Javascript.²³

Metadati. Dati che descrivono e definiscono altri dati in un determinato contesto. Il contesto riguarda le condizioni in cui avviene il trattamento dei dati. Meta è un prefisso che nel linguaggio informatico indica una definizione o descrizione. Quindi metadati è una definizione o una descrizione di dati, così come i metatag indicano quei tag inseriti in una pagina che ne descrivono i contenuti. I metadati arricchiscono il contenuto informativo dei dati esplicitandone delle proprietà che semplificano il processo di fruizione dei dati stessi, facilitandone, la ricerca, il recupero, la composizione e di conseguenza il riutilizzo.²⁴

Namespace. Contenitore di nomi di entità, raggruppamento logico di identificatori univoci o di simboli.²⁵ Namespace dichia-

²³ Per una più ampia spiegazione del meccanismo del mashup vedi: <[http://it.wikipedia.org/wiki/Mash-up_\(informatica\)](http://it.wikipedia.org/wiki/Mash-up_(informatica))>

²⁴ <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

²⁵ Per un approfondimento vedi: Mirna Willer, Gordon Dunsire, Boris Bosančić, *ISBD and the semantic web*, "Jlis.it", vol. 1, no. 2 (dicembre/December 2010), p. 213-236; vedi inoltre: <<http://www.gordondunsire.com/presentations.htm>>.

ra un ambito che contiene un insieme di oggetti correlati, un raggruppamento di nomi sotto un unico identificatore. È possibile utilizzare uno spazio dei nomi per organizzare gli elementi di codice e creare tipi globalmente univoci. Con il termine namespace ci si riferisce, quindi, a un insieme di nomi usati come tipi di elemento (tag) e attributi di questi elementi. In un documento/file XML possono essere usati elementi (e loro attribuiti) diversi, qualificati con un prefisso (per esempio DC, FOAF, SKOS, etc.), in modo da evitare confusioni. Il prefisso viene definito richiamando il namespace nel documento XML mediante l'attributo `xmlns`. Per esempio, `<bibliografia xmlns:dc="http://purl.org/dc/elements/1.1/">` specificato nella radice di un documento XML permette di usare al suo interno, in ogni posizione, tag definiti in Dublin Core (DC) quali `dc:creator`, `dc:title`, `dc:rights`, etc. Il namespace è, dunque, una collezione di tutti gli URI che utilizzano lo stesso dominio. Gli URI `http://iflastandards.info/ns/isbd/terms/mediatype/T1007` e `http://iflastandards.info/ns/isbd/terms/mediatype/T1003` sono gli URI dei concetti video e microforma presi dal namespace definito per il vocabolario dei tipi di media di ISBD, che ha come dominio di base `http://iflastandards.info/ns/isbd/terms/mediatype`. Il vantaggio nell'uso di namespace è che il dominio di base può essere abbreviato per semplificare la comprensione umana dell'URI: negli esempi, il dominio di base può essere abbreviato con 'isbdmedia'; questa parte abbreviata dell'URI è definita QName e può essere congiunta con la parte locale dell'URI (/T1007 e /T003 negli esempi) con due punti (:), per formare un *URI compatto* o *CURIE*: `isbdmedia:T007` e `isbdmedia:T003`. Il QName è automaticamente sciolto nella forma estesa quando letto dalle macchine.

Ontologia. Alla base della gestione della conoscenza c'è l'ontologia che permette una rappresentazione formale della conoscenza di un particolare dominio, tramite la definizione dei concetti, degli oggetti, delle loro proprietà e delle regole logiche di relazione tra di essi. Nell'Information Technology il termine ontologia viene utilizzato per rappresentare il sapere inscritto in un certo dominio che può essere sia una realtà dove tutta la conoscenza è a disposizione degli esperti del dominio, sia una realtà dove una parte della conoscenza è esplicitata, per esempio,

in una base documentale, oppure un altro ambito in cui i legami logici tra le informazioni si deducono dalla loro strutturazione in una banca dati. A seconda delle diverse esigenze, sono stati sviluppati alcuni linguaggi standard (RDF, F-Logic, OWL) che hanno lo scopo di rendere universale la rappresentazione e la gestione delle ontologie. Scegliendo di rappresentare un'ontologia si adotta, quindi, contemporaneamente uno standard di comunicazione condiviso da diversi sistemi informatici e applicativi.²⁶

Open data. Dati pubblicati con licenze open e, dunque, liberamente disponibili per chiunque voglia utilizzarli o ripubblicarli. Nel concetto di *open* è inclusa la componente relativa alla possibilità di (ri)utilizzo e, pertanto, la necessità che siano resi disponibili tramite formati aperti standardizzati che rendano i dati effettivamente utilizzabili. Gli aspetti più importanti degli open data sono: disponibilità e accesso, riutilizzo e redistribuzione, partecipazione universale.²⁷

Open Metadata Registry. Nel contesto biblioteconomico, l'Open Metadata Registry (OMR), <<http://metadataregistry.org/>>, fornisce gli strumenti per identificare, dichiarare e pubblicare schemi di metadati (elementi e proprietà), vocabolari controllati e profili applicativi (AP). L'Open Metadata Registry supporta, inoltre, il mapping di relazioni tra termini e concetti appartenenti a vocabolari e schemi differenti. Suoi obiettivi principali sono la scoperta, il riuso, la standardizzazione e l'interoperabilità dei metadati a livello locale e globale.

Open source. Insieme di software disponibile al pubblico in forma di codice sorgente e che non ha restrizioni di licenza che ne limita l'uso, la modifica e la redistribuzione. Il software open source è in genere coperto da copyright, e la sua licenza può contenere restrizioni intese a preservare il suo stato di open source, notizie sull'autore, oppure controlli di sviluppo. Il termine open source è stato registrato come marchio di certificazione dalla

²⁶ <<http://www.cineca.it/it/content/knowledge-management>>.

²⁷ <Open Data Handbook, <<http://opendatahandbook.org/it/what-is-open-data/>>.

Software in the Public Interest, la quale ha anche sviluppato la definizione ufficiale di open source.²⁸

OWL (Web Ontology Language). Linguaggio ontologico progettato per fornire un linguaggio che può essere usato per descrivere le classi e le relazioni che le collegano e che sono inerenti ai documenti e alle applicazioni web.²⁹

N-Triples. Serializzazione di RDF in cui ogni tripla è espressa interamente (i.e., non in modo sintetico) e indipendentemente dalle altre. La concatenazione delle triple di un dataset RDF secondo N-Triples avviene mediante il carattere punto (i.e., <soggetto1><predicato1><oggetto1>.<soggetto2><predicato2><oggetto2>).³⁰

Proof e TRUST. Livelli di controllo e validazione applicati dal sistema per verificare che i dati provengano da fonti affidabili (trusted), i quali, uniti all'applicazione dei principi della logica formale per elaborare i dati derivati, consentono di fornire risultati univoci e affidabili, garantendo un elevato livello di attendibilità agli utilizzatori. TRUST si presenta come uno strumento fondamentale perché consente la gestione di informazioni codificate digitali non più familiari.³¹

R2RML (Relational to RDF Mapping Language). Linguaggio per la trasformazione di basi dati relazionali in RDF, con possibilità di riutilizzare vocabolari e ontologie esistenti durante il processo di conversione. Alcuni tool di esempio: Triplify, D2R Server, Virtuoso RDF ViewsWeb.

RDF (Resource Description Framework). Strumento base proposto da W3C per la codifica, lo scambio e il riutilizzo di

²⁸ <<http://opensource.org/osd.html>>.

²⁹ <<http://www.w3.org/TR/2004/REC-owl-guide-20040210/>>.

³⁰ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_publicolg2014_v0.7finale.pdf>.

³¹ <<http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/09/APARSEN-Trust-Brochure-Low-Res-Web-Version.pdf>>.

metadati strutturati, consente l'interoperabilità tra applicazioni che condividono le informazioni sul web. È costituito da due componenti: (1) RDF Model and Syntax: espone la struttura del modello RDF, e descrive una possibile sintassi; (2) RDF Schema: espone la sintassi per definire schemi e vocabolari per i metadati. RDF Data Model si basa su tre principi chiave: (1) Qualunque cosa può essere identificata da un URI (Universal Resource Identifier); (2) The least power: utilizzare il linguaggio meno espressivo per definire qualunque cosa; (3) Qualunque cosa può dire qualunque cosa su qualunque cosa. RDF consente di catturare la semantica dei dati, quindi la loro comprensibilità, e di facilitarne l'accessibilità da parte di agenti automatici tramite l'infrastruttura e i protocolli internet esistenti.

RDFa. Specifica una sintassi tramite attributi che consentono di esprimere dati strutturati in pagine HTML; non definisce un vocabolario di termini specifici, ma utilizza specifiche indipendenti di termini definite altrove (vocabolari e tassonomie).

RDF crawling. Motori di ricerca del web semantico utilizzano robot per recuperare in modo automatico dati RDF dai contenuti del web, permettendone così la ricerca e la navigazione sui dati.³²

RDFS (Resource Description Framework Schema). Sintassi RDF per definire schemi di dati e vocabolari per i metadati; in particolare, fornisce un vocabolario di data-modelling per dati RDF. In RDF ogni predicato è astratto, senza relazioni con altri predicati; in RDFS è, invece, possibile esprimere relazioni e vincoli tra predicati e definire le proprietà di un concetto.

Record. Unità di una struttura dati distinta in campi, ciascuno dei quali ha un nome e un tipo di dato propri. Nella formalizzazione delle teorie delle basi di dati relazionali il termine relazione conserva il suo significato matematico: considerati n domini (non necessariamente distinti) $D_1, D_2, \dots, D_n$, una

³² <www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/SemanticWebSearchEngines>

relazione R di questi n domini è un insieme di n -tuple, in ciascuna delle quali il primo elemento appartiene a D_1 , il secondo a D_2 e così via; ovvero R è un sottoinsieme del prodotto cartesiano degli n domini $D_1 \times D_2 \times \dots \times D_n$. I domini hanno tipi di dato diversi, per esempio, numeri interi, numeri in virgola mobile, booleani (valore 0 o 1), sequenze di caratteri alfabetici (stringhe) a lunghezza fissa o variabile. In modo concreto una relazione può essere rappresentata come una tabella a cui è associato un nome che la identifica, ogni riga rappresenta una n -tupla della relazione, l'ordine delle righe non è significativo, tutte le righe sono diverse tra loro e corrispondono ai domini in cui la relazione è definita e anche a loro viene attribuito un nome distinto. Secondo una terminologia più tradizionale ogni riga della tabella viene detta *record* e ciascuna colonna della riga è denominata *campo*. I campi contengono i valori cioè i veri e propri dati del record.

REST (Representational State Transfer). Astrazione degli elementi architetturali di un sistema distribuito ipermediale che ignora i dettagli relativi all'implementazione e alla sintassi dei protocolli, per focalizzarsi sui ruoli dei componenti, i vincoli della loro interazione con altri componenti e la loro interpretazione degli elementi di dati significativi. I concetti chiave di REST sono la *risorsa*, ogni informazione alla quale può essere dato un nome e i *componenti*. Le risorse possiedono *identificatori* e hanno *rappresentazioni*. I componenti REST compiono azioni su una risorsa usando una rappresentazione per catturarne lo stato attuale o previsto e trasferendo quella rappresentazione tra i componenti. Una rappresentazione è una sequenza di byte (sequenza binaria di 8 bit), più i metadati che li descrivono.³³

RIF (Rules Interchange Format). Linguaggio XML raccomandato dal W3C per esprimere regole che un computer possa leggere ed eseguire. È un linguaggio che serve per istruire le macchine.

³³ Roy Thomas Fielding, *Architectural styles and the design of network-based software architectures. Doctoral Thesis Dissertation*, University of California, Irvine, 2000, <<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>>.

Riuso di dati. Qualsiasi uso dei dati diverso da quello per il quale sono stati prodotti o raccolti originariamente.³⁴ Il Codice per l'Amministrazione digitale prevede³⁵ che le pubbliche amministrazioni prima di procedere all'acquisizione o allo sviluppo di sistemi informatici provvedano all'analisi comparativa con soluzioni già sviluppate e disponibili per il riuso. Inoltre l'art. 69 prescrive al comma 1 che "Le pubbliche amministrazioni che siano titolari di programmi informatici realizzati su specifiche indicazioni del committente pubblico, hanno obbligo di darli in formato sorgente, completi della documentazione disponibile, in uso gratuito ad altre pubbliche amministrazioni che li richiedono e che intendano adattarli alle proprie esigenze, salvo motivate ragioni".³⁶

SOAP (Simple Object Access Protocol). Protocollo che consente lo scambio di informazioni strutturate tramite messaggi in un ambiente decentralizzato, distribuito in rete ed eterogeneo, avvalendosi di diversi protocolli sottostanti in particolare HTTP.

Schema.org. Progetto nato dalla collaborazione tra Google, Microsoft e Yahoo! con lo scopo di migliorare il web creando uno schema di markup dei dati strutturati supportato dai principali motori di ricerca. Il markup all'interno della pagina consente ai motori di ricerca di comprendere le informazioni presenti nelle pagine web e fornire risultati più completi. Una terminologia di markup condivisa consente ai webmaster di decidere più facilmente quale schema di markup utilizzare e di ottenere i massimi vantaggi dal loro lavoro.³⁷

Serializzazione: Processo di traduzione di strutture di dati o stato di un oggetto in un formato che può essere memorizzato (in un file o trasmesso in rete) e ricostruito più tardi (deserializzazione), nello stesso computer o in un ambiente diverso.

³⁴ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicolg2014_v0.7finale.pdf>.

³⁵ <<http://www.agid.gov.it/agenda-digitale/cad/capo-vi-sviluppo-acquisizione-riuso-sistemi-informatici-nelle-pubbliche>>.

³⁶ <<http://www.agid.gov.it/cad/riuso-programmi-informatici>>.

³⁷ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicolg2014_v0.7finale.pdf>.

Soggettari. Strutture che organizzano i concetti tramite relazioni di tipo gerarchico, usualmente basati su insiemi di termini utilizzabili per formulare il soggetto, inteso come “il tema di base”, di un’opera, tramite due relazioni: equivalenza e gerarchico associativa.

SPARQL endpoint. Standard W3C per eseguire interrogazione dati RDF e per recuperare i dati da un triple store RDF (archivio come un insieme di triple soggetto – predicato – oggetto), per ottenere le risposte con le query complesse e di eseguire diverse analisi multidimensionali sui dati. Molti SPARQL endpoint sono pubblicamente disponibili e diversi soggetti, produttori e gestori dei dati, hanno costruito SPARQL Endpoint per accedere ai dati aperti.

Sussunzione. L’atto di ricondurre un concetto nell’ambito di uno più ampio che lo comprenda. “L’atto e il procedimento del sussumere, cioè del ricondurre un concetto nell’ambito di quello nella cui estensione esso è compreso: *giudizi di s.*, o *sussumtivi* (per esempio «il cane è un mammifero»); *sillogismi di s.*, o *sussuntivi*, quelli di prima figura, in cui la gerarchia estensiva dei termini appare più direttamente ed esclusivamente messa in luce (per esempio, «il cane è un mammifero; i mammiferi sono vertebrati, dunque il cane è un vertebrato»)”.³⁸

SKOS (Simple Knowledge Organization System). Area di lavoro che sviluppa specifiche e standard per supportare l’uso dei sistemi d’organizzazione della conoscenza (KOS) nel quadro del web semantico. SKOS Core Vocabulary è stato progettato con l’idea di essere facilmente estensibile: esso vuole fare da connettore tra i più tradizionali sistemi di organizzazione della conoscenza usati nelle biblioteche, musei, archivi come i thesauri e i sistemi di classificazione e le nuove strutture pensate per il Web come le open-directory, i blog topic exchange. SKOS Core Vocabulary consiste in una serie di classi RDFS e proprietà di RDF che sono usate per rappresentare il contenuto e la struttura di base dei cosiddetti concept-schema cioè thesaurus, schemi

³⁸ <www.treccani.it/vocabolario/sussunzione/>.

di classificazione, elenchi di titoli d'argomento, tassonomie, terminologie, glossari e altri tipi di vocabolario.³⁹

SPARQL (Sparql Protocol And Rdf Query Language). Linguaggio standardizzato dal W3C con una sintassi simile a quella SQL per l'interrogazione di dati RDF e protocollo di comunicazione basato su HTTP. Una semplice interrogazione SPARQL si compone di una concatenazione di triple in cui alcuni elementi possono essere delle variabili incognite. L'esecuzione di una query SPARQL cerca tra i dati le concatenazioni di triple "conformi" a quelle della query, assegnando (i.e., istanziando) degli URI o dei valori alle variabili che possono anche essere restituiti in output. Inoltre, in un'interrogazione è possibile specificare operatori più complessi come disgiunzioni, triple opzionali, filtri, espressioni regolari, etc. I servizi web che implementano il protocollo SPARQL sono detti SPARQL endpoint.⁴⁰

Standard aperto – Open Standard. La Free Software Foundation lo definisce come un formato o un protocollo soggetto a un processo di valutazione (assessment) pubblico, privo di componenti o estensioni dipendenti da altri standard che non soddisfino la definizione di standard aperto; è libero da clausole tecniche e legali che ne limitino l'utilizzo da parte di chiunque e in ogni contesto; è gestito e sviluppato in modo indipendente da ogni singola azienda, secondo un processo aperto alla partecipazione dei vari collaboratori; è disponibile in molteplici implementazioni. Gran parte degli standard del web (HTTP, URI, HTML, RDF) sono standard aperti.

Tag. Contrassegno mediante il quale, in un testo memorizzato, si individuano singole entità, rendendole così riconoscibili dal mezzo elettronico ed estraibili con procedimento automati-

³⁹ <<http://www.iskoi.org/doc/skos.htm>>; <<http://www.w3.org/2004/02/skos/core#>>.

⁴⁰ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>; <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopen_data_v2.0_0.pdf>; in Italia i dati del Senato della Repubblica sono interrogabili tramite SPARQL endpoint; vedi: <<http://dati.senato.it/23>>.

co.⁴¹ Nei linguaggi di marcatura come HTML e XML, i tag sono i nomi (etichette) racchiusi tra i simboli “<” e “>” (*start-tag*) o tra i simboli “<” e “/>” (*end-tag*) che permettono di marcare gli elementi (per esempio <title>.....</title>, <h1>.....</h1>, <dc:creator>....</dc:creator>) o di includere un contenuto esterno (*empty-tag*, per esempio <img...>).

Tassonomia. Classificazione gerarchica di concetti ed elementi che consente di rappresentare un dominio di conoscenza.⁴² Schemi di categorie o sottocategorie che possono essere impiegate per raggruppare o per organizzare elementi di informazione e conoscenza. Le tassonomie variano per tipologia, da semplici a complesse. In quelle complesse le categorie possono essere suddivise in sottocategorie, con più livelli gerarchici.⁴³

Thesauri. Collezione di termini privi di definizione. Thesaurus e vocabolario sono utilizzati in maniera pressoché equivalente.⁴⁴ I thesauri sono vocabolari controllati e strutturati in cui i concetti sono rappresentati da termini, organizzati in modo che le relazioni tra concetti siano esplicite e i termini preferiti siano corredati da sinonimi e quasi sinonimi. Scopo di un thesaurus è guidare l'indicizzatore e il ricercatore a scegliere lo stesso termine preferito o la combinazione di termini preferiti per presentare un concetto. Per questa ragione un thesaurus è ottimizzato per la navigabilità umana e per la copertura terminologica di un dominio. I thesauri sono basati su una struttura gerarchica a tre relazioni distinte: equivalenza, gerarchica e associativa.⁴⁵

⁴¹ <http://dizionari.corriere.it/dizionario_italiano/T/tag.shtml>.

⁴² <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_publicolg2014_v0.7finale.pdf>.

⁴³ Vedi la traduzione italiana della voce *Taxonomy* nel glossario dello standard: International Organization for Standardization, ISO 25964-2: *Information and documentation -thesauri and interoperability with other vocabularies – Part. 2: Interoperability with other vocabularies*. Geneva: International Organization for Standardization, 2013, p. 14.

⁴⁴ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_publicolg2014_v0.7finale.pdf>.

⁴⁵ Vedi la traduzione italiana della voce *Thesaurus* nel glossario dello standard: International Organization for Standardization, ISO 25964-2: *Information and documentation -thesauri and interoperability with other*

Tripla. Il web semantico ha un suo linguaggio con una sua grammatica, funzionale a esprimere la natura relazionale dei dati e la loro proteiforme tipologia. Questa grammatica è RDF (Resource Description Framework); essa fornisce la struttura logica per gestire ed esprimere le relazioni tra le informazioni, ispirandosi ai principi della logica dei predicati, secondo la quale le informazioni sono esprimibili mediante *asserzioni* (statement), costituite da un modello sintagmatico tripartito, chiamato *tripla*, che contempla la sequenza di tre elementi: soggetto, predicato e oggetto.

1. *soggetto*, una qualsiasi *risorsa*, che identifica l'oggetto descritto;
2. *predicato*, una *proprietà* specifica della risorsa o una relazione utilizzata per descriverla, identificata da un URI;
3. *oggetto*, detto *valore*.

Il ruolo assolto da ciascun elemento riflette la struttura logica del linguaggio naturale. La sequenza dei tre elementi è definita asserzione (*statement*). Il soggetto e il predicato devono essere identificati tramite un URI, mentre l'oggetto può essere sia un URI, sia un dato nominale.

Tupla RDF. Un set di asserzioni descritte in triple. Nella teoria delle basi di dati, una tupla è un elemento generico di una relazione con attributi in un database relazionale. Si tratta di una riga di una tabella o relazione, cioè dell'insieme dei valori assunti dai campi "dato" o "attributi" specificati in cima a ciascuna colonna. Essa rappresenta, dunque, l'entità base che dev'essere memorizzata nella base di dati, identificabile con il record. Una tupla RDF è differente da una tripla RDF perché mentre la tripla RDF è sempre composta da tre parti, identificate da nomi fissi (soggetto, predicato e oggetto), una tupla RDF può avere un numero vario di componenti, senza un nome definito.

Unicode. Sistema di codifica che assegna un numero univoco a ciascun carattere usato per la scrittura di testi, in maniera indipendente dalla lingua, dalla piattaforma informatica e dal programma utilizzato.⁴⁶

vocabularies – Part. 2: Interoperability with other vocabularies. Geneva: International Organization for Standardization, 2013, p. 14.

⁴⁶ Da *Wikipedia*.

Unifying logic. Logica di unificazione del sistema.

URI (Uniform Resource Identifier). Stringa di caratteri che identifica univocamente una risorsa: pagina web, documento, immagine, file, etc.⁴⁷

URL (Uniform Resource Locator). Stringa di caratteri che identifica una risorsa su internet, ne specifica formalmente la localizzazione e indica il protocollo da utilizzare per accedervi. È un tipo specifico di URI.⁴⁸

User interface. Strato finale che permetterà agli esseri umani di utilizzare applicazioni del web semantico.

VIAF (Virtual International Authority File). Progetto internazionale che ha dato vita a un'unica base dati di voci d'autorità controllate, (circa 20 milioni di autori provenienti da 24 cataloghi collettivi nazionali) a disposizione quale servizio gratuito del web semantico; in essa l'utente può cercare un autore utilizzando la forma del nome nella propria lingua, scrittura e alfabeto preferiti e accedere immediatamente ai cataloghi nazionali o internazionali corrispondenti per visualizzare la lista delle loro opere.⁴⁹

URI utilizzati per il record della risorsa *Dante Alighieri*:

```
<rdf:Description rdf:about="http://viaf.org/viaf/97105654">
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
<rdf:type rdf:resource="http://rdvocab.info/uri/schema/
FRBRentitiesRDA/Person"/>
<foaf:name>Dante Alighieri, 1265-1321</foaf:name>
<owl:sameAs rdf:resource="http://d-nb.info/gnd/118523708"/>
<owl:sameAs rdf:resource="http://libris.kb.se/resource/auth/
183027"/>
<owl:sameAs rdf:resource="http://data.bnf.fr/ark:/12148/
cb118985852#foaf:Person"/>
```

⁴⁷ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_publicolg2014_v0.7finale.pdf>.

⁴⁸ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_publicolg2014_v0.7finale.pdf>.

⁴⁹ <http://it.wikipedia.org/wiki/Virtual_International_Authority_File>.

```

<owl:sameAsrdf:resource="http://www.idref.fr/026811081/
id"/>
<owl:sameAsrdf:resource="http://dbpedia.org/resource/
Dante_Alighieri"/>
</rdf:Description>

```

VOAF (Vocabulary of a Friend). Vocabolario che consente di descrivere i termini usati impiegando gli specifici linked data. Esso consente di mettere in relazione diversi vocabolari o di indicare quali dataset ne fanno uso.⁵⁰

Vocabolario. Insieme dei possibili valori che le entità di una classe possono assumere all'interno di un dominio di conoscenza o di una ontologia.⁵¹

VoID (Vocabulary of Interlinked Datasets). Vocabolario che consente di rappresentare metadati sui dataset RDF, in particolare, per caratterizzare le informazioni che riguardano le connessioni tra diversi dataset, come, per esempio, i linkset.⁵² Con VoID la scoperta e l'uso dei dataset con link possono essere rappresentati effettivamente ed efficacemente.⁵³

W3C Library Linked Data Incubator Group. Gruppo di lavoro internazionale costituito in sede W3C il cui scopo è facilitare lo sviluppo dell'interoperabilità globale dei dati delle biblioteche nel web; riunire le persone coinvolte nelle attività del web semantico, focalizzate sui linked data; creare iniziative e modalità di collaborazione per il futuro. Il gruppo condesidera suo impegno definire i requisiti per gli standard e le linee guida nel web.⁵⁴

⁵⁰ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

⁵¹ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

⁵² <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

⁵³ <<http://semanticweb.org/wiki/VoID>>.

⁵⁴ <<http://www.w3.org/2005/Incubator/ldd/>>.

Web semantico. Insieme di modelli e standard web in cui le risorse vengono descritte e correlate tra loro in modo formale tramite l'uso opportuno di metadati. In questo modo si abilitano gli agenti automatici a comprendere il significato dei dati e delle informazioni.⁵⁵

Web services e Web services RDF. Componente software che gestisce l'interazione tra applicazioni collegate, di solito su sistemi diversi e distribuiti in rete. Insieme alla sua programmazione (tecnicamente *implemetazione*) viene specificata anche la sua interfaccia (descritta in formato XML, specificatamente in linguaggio WSDL-Web Services Description Language), che descrive il comportamento del servizio: come richiamarlo, quali risultati può restituire, in quale formato è disponibile, quali errori può generare, quali protocolli di rete supporta. La descrizione – che essendo in XML, può essere elaborata automaticamente – favorisce lo sviluppo delle applicazioni che per usare il web service devono attenersi alle modalità prescritte nella sua descrizione, usando messaggi conformi alla specifica SOAP del W3C, trasmessi mediante il protocollo HTTP e serializzati in XML, avvalendosi di altri standard web.⁵⁶ In tempi più recenti sono stati definiti anche web services di tipo REST che usano sempre HTTP e XML, ma che sono definiti secondo un modello diverso rispetto a quello SOAP e sono basati sui concetti di risorse, rappresentazione delle risorse e trasferimento delle rappresentazioni. Gli RDF web services sono dei web services che permettono di far interoperare tra loro i servizi dello stack applicativo RDF distribuiti in rete su sistemi diversi.

XML (eXtensible Markup Language). Linguaggio di marcatura standardizzato proposto dal W3C usato per la descrizione di documenti e per la costruzione di altri linguaggi più specifici (per esempio, XBRL per la rappresentazione dei bilanci, Normattiva per la rappresentazione di documenti informatici in

⁵⁵ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicolg2014_v0.7finale.pdf>.

⁵⁶ *Web Services Glossary W3C Working Group Note 11 February 2004*, <<http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/>>.

ambito giuridico, etc.). Esistono strumenti che possono essere utilizzati per definire la struttura dell'XML, come DTD e XML schema. Il mondo legato all'XML è estremamente ampio; una serializzazione (formato fisico) di RDF fa uso di sintassi XML.⁵⁷ XML si è consolidato come lingua franca, ovvero come standard per garantire l'interoperabilità in rete sia per i messaggi dei protocolli e dei servizi usati dalle applicazioni per interagire tra loro, sia per la strutturazione dei dati.

XML Schema o Schema XML. Linguaggio di descrizione del contenuto di un file XML, il primo – e al momento unico – che abbia ottenuto la validazione ufficiale del W3C (col valore 1.1); suo scopo è delineare gli elementi permessi, i tipi di dato associati e le relazioni gerarchiche tra gli elementi in un file XML. L'XML Schema, in sintesi, consente:

- a) la validazione di file/documenti XML ben strutturati;
- b) l'interoperabilità tra sistemi diversi che sono così in grado di interpretare il documento.

XML namespaces. Garantisce la possibilità di utilizzare dati provenienti da più fonti. Una caratteristica precipua del web semantico è la capacità di creare relazioni tra fonti, identificando i dati in modo inequivocabile.

⁵⁷ <http://www.agid.gov.it/sites/default/files/linee_guida/patrimonio_pubblicol2014_v0.7finale.pdf>.

RIFERIMENTI BIBLIOGRAFICI

Gli URL sono stati controllati l'ultima volta il 26 gennaio 2015.

- @Cult blog, *Biblioteche e linked data*, 2011, <<https://atcult.wordpress.com/2011/01/14/biblioteche-e-linked-data/>>.
- Sören Auer, Volha Bryl, Sebastian Tramp, *Criating knowledge out of interlinked data. Results of the LOD2 Project*, Springer Open, Lecture notes in Computer Science, 2014, <<http://link.springer.com/book/10.1007%2F978-3-319-09846-3>>.
- Tomas Baker, *Designing data for the open world of the web*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/6308/7866>>, p. 1.
- Richard Cyganiak, Anja Jentzsch, *Linking Open Data cloud diagram*, <<http://lod-cloud.net/>>.
- D2RQ – *Accessing relational databases as virtual RDF graphs*, <<http://d2rq.org/>>.
- Giovanni Bergamin, Anna Lucarelli, *The Nuovo Soggettario as a service for the linked data world*, "Jlis.it.", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cineca.it/index.php/jlis/article/view/5474/7904>>.
- Jesús Barrasa, Óscar Corcho, Asunción Gómez-Pérez, *R2O an extensible and semantically based database-to-ontology mapping language*, <http://www.cs.man.ac.uk/~ocorcho/documents/SWDB2004_BarrasaEtAl.pdf>.
- Tim Berners-Lee, *L'architettura del nuovo web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*; in collaborazione con Mark Fischetti. Milano: Feltrinelli, 2001. Traduzione di: *Weaving the web: the past, present and future of the world wide web by its inventor*; with Mark Fischetti. London: Orion Business Book, 1999.
- Tim Berners-Lee, *Linked data - Design issues*, 2006, <<http://www.w3.org/DesignIssues/LinkedData.html>>.
- Tim Berners-Lee, Christian Bizer, Tom Heath, *Linked data: the story so far*, <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>.
- Tim Berners-Lee, James Hendler, Ora Lassila, *The semantic web. A new form of web content that is meaningful to computers will unleash a*

- revolution of new possibilities*, “Scientific American magazine”, May 2001, <<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>>.
- Tim Berners-Lee, *Semantic web road map*, 1998, <<http://www.w3.org/DesignIssues/Semantic.html>>.
- Tim Berners-Lee, *Universal Resource Identifiers – Axioms of web architecture*, 1996, <<http://www.w3.org/DesignIssues/Axioms.html>>.
- Carlo Bianchini, *Dagli OPAC ai library linked data: come cambiano le risposte ai bisogni degli utenti*, “AIB studi”, vol. 52, n. 3 (settembre/dicembre 2012), p. 303-323.
- Carlo Bianchini, *Il falso in internet: autorevolezza del web, information literacy e futuro della professione (seconda parte)*, “AIB studi”, vol. 54, n. 2/3 (maggio/dicembre 2014), p. 227-240.
- Carlo Bianchini, *Futuri scenari: RDA, REICAT e la granularità dei cataloghi*, “Bollettino AIB”, vol. 50, n. 3 (settembre 2010), p. 219-238.
- Carlo Bianchini, *Organizzare la conoscenza con la sequenza di filiazione della Classificazione Colon di S. R. Ranganathan*, “Jlis.it”, vol. 2, n. 2 (dicembre/December 2011), p. 1-21.
- Carlo Bianchini, *FRBR prima di FRBR. Il numero di libro nella Colon Classification*, “Jlis.it”, vol. 1, no. 1 (gennaio/January 2010), p. 11-39.
- Carlo Bianchini, *Rassegna critica dei recenti sviluppi della catalogazione in ambito internazionale*, “Bollettino AIB”, vol. 49, n. 1 (gennaio 2009), p. 35-44.
- Carlo Bianchini, *Riflessioni sull’universo bibliografico. Funzioni, oggetti e modelli della catalogazione per autore e per titolo*, prefazione di Mauro Guerrini. Milano: Sylvestre Bonnard, 2005.
- Carlo Bianchini, Mauro Guerrini, *Introduzione a RDA. Linee guida per rappresentare e scoprire le risorse*; prefazione di Barbara B. Tillett; postfazione di Gordon Dunsire. Milano: Editrice Bibliografica, 2014.
- Carlo Bianchini, Mauro Guerrini, *RDA: Resource Description and Access. Il nuovo standard per la metadattazione e la scoperta delle risorse nell’era digitale*, “Biblioteche oggi”, vol. 32, n. 4 (maggio 2014), p. 46-50.
- Carlo Bianchini, Mauro Guerrini, *RDA: Resource Description and Access. The new standard for metadata and resource discovery in the digital age*, “Jlis.it”, vol. 6, no. 1 (January 2015), p. 21-31, <<http://leo.cineca.it/index.php/jlis/article/view/10963/10150>>.

- Carlo Bianchini, Mauro Guerrini, *Quis custodiet ipsos custodies? Osservazioni sulle relazioni fra FRBR, ICP, ISBD e RDA*, "Il bibliotecario", 3 serie, n. 1 (gennaio–aprile 2009), p. 59-85.
- Carlo Bianchini, Mirna Willer, *ISBD resource and its description in the context of the semantic web*, "Cataloging & Classification Quarterly", vol. 52, no. 8 (2014), p. 869-887.
- Attilio Mauro Caproni, *Bibliografia generale*. Napoli: Società Editrice Napoletana, 1980.
- Karen Coyle, *Library data in a modern context*, "ALA Techsource", vol. 46, no. 1 (2010), p. 5-13, <<http://alatechsource.metapress.com/content/p3022442071g7655/fulltext.pdf>>.
- Karen Coyle, *Linked data: an evolution*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), p. 53-61, <<http://leo.cilea.it/index.php/jlis/article/view/5443/>>.
- Karen Coyle, *Linked data tools: connecting on the web*, report ALA, "Library Technology Reports", vol. 48, no. 4 (May/June 2012), <<https://www.questia.com/library/journal/1G1-292236286/resources>>.
- Karen Coyle, *Nuove direzioni in metadata* [video], 2010, <<http://www.archive.org/details/NuoveDirezioniInMetadata>>.
- Karen Coyle, *Understanding the semantic web: bibliographic data and metadata*, report ALA "Library Technology Reports", vol. 46, no. 1 (January 2010), <<https://www.questia.com/library/journal/1G1-292236286/resources>>.
- Karen Coyle *on the web, linked data links – Contents*, <<http://kcoyle.net/presentations/links.html>>.
- Gianfranco Crupi, *Oltre le Colonne d'Ercole: linked data e cultural heritage*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/8587>>.
- Gianfranco Crupi, *Universo bibliografico e semantic web*, Digilab, <https://www.academia.edu/4031499/Universo_bibliografico_e_semantic_web>.
- Alan Danskin, *Linked and open data: RDA and bibliographic control*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/5463>>.
- Isa De Pinedo, Alberto Petrucciani, *Un approccio all'applicazione del modello FRBR alle regole di catalogazione italiane: problemi e possibili soluzioni*, "Bollettino AIB", vol. 42, n. 3 (settembre 2002), p. 267-278.

- Antonella De Robbio, *Forme e gradi di apertura dei dati*, “Biblioteche oggi”, vol. 30, n. 6 (luglio–agosto 2012), p. 11-24.
- Antonella De Robbio, Silvia Giacomazzi, *Dati aperti con LODE*, “Bibliotime”, a. 14, n. 2 (luglio 2011), <<http://www.aib.it/aib/sezioni/emr/bibtime/num-xiv-2/derobbio.htm>>.
- Stefan Decker, Michael Hausenblas, Jodi Schneider, Nuno Lopes, *Linked data in the digital humanities*, “Lodlam.net”, 2012, <<http://lodlam.net/2012/10/27/linked-data-in-the-digital-humanities/>>.
- Roberto Delle Donne, *Un intreccio di iniziative scientifiche. Reti medievali e il futuro della storiografia digitale*, “Reti medievali rivista”, vol. 15, n. 2 (dicembre 2014), p. 93-156, <<http://rivista.retimedievali.it>>.
- Dereferencing HTTP URIs <<http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14>>.
- Gordon Dunsire, *The role of ISBD in the linked data environment*, “Cataloging & Classification Quarterly”, vol. 52, no. 8 (2014), p. 855-868.
- Gordon Dunsire, *UNIMARC and linked data*, “IFLA Journal”, vol. 37, no. 4 (December 2011), p. 314-326.
- Gordon Dunsire, Diane Hillmann, Jon Phipps, *Reconsidering universal bibliographic control in light of the semantic web*, “Journal of Library Metadata”, vol. 12, nos. 2/3 (2012), p. 164-176.
- Gordon Dunsire, Mirna Willer, *Standard library metadata models and structures for the semantic web*, “Library Hi Tech News”, vol. 28, no. 3 (2011), p. 1-12.
- Elena Escolano Rodríguez, *ISBD adaptation to SW of bibliographic data in linked data*, “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/5484>>.
- Elena Escolano Rodríguez, *ISBD en la web semántica : lectio magistralis en Biblioteconomía : Florencia, Italia, Universidad de Florencia, 6 de marzo de 2012 = ISBD nel web semantico : lectio magistralis in Biblioteconomía : Firenze, Italia, Università di Firenze, 6 marzo 2012*. Fiesole (Firenze): Casalini Libri, 2012, <<http://store.torrossa.it/resources/an/2488606?ref=http://leo.cineca.it/index.php/jlis/pages/view/lectio>>..
- The Europeana Data Model: constraints and opportunities*. Europeana Connect, Prof. Dr. Stefan Gradmann based on work done with Martin Doerr, Hennicke Steffen, Antoine Isaac, Carlo Meghini, Guus Schreiber, Herbert Van de Sompel, slide 6, <<http://www>>.

- europeanaconnect.eu/europeanatech/documents/Gradmann_Stefan_LinkedOpenEuropeana.pdf>.
- Roy Thomas Fielding, *Architectural styles and the design of network-based software architectures*. Doctoral Thesis Dissertation, University of California, Irvine, USA, 2000, <<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>>.
- Giuliano Genetasio, *The International Cataloguing Principles and their future*, "Jlis.it", vol. 3, no. 1 (gennaio/January 2012), p. 1-17.
- Global Interoperability and Linked Data in Libraries*; special issue, editor: Mauro Guerrini, proceedings by Gianfranco Crupi and Ginevra Peruginelli, "Jlis.it", vol. 4, no.1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/issue/view/536>>.
- Mauro Guerrini, *BIBFRAME. Per un nuovo ruolo delle biblioteche nel contesto del web*, "Digitalia. La rivista del digitale nei beni culturali", vol. 1 (giugno 2014), p. 125-126. <<http://digitalia.sbn.it/article/view/1061/691>>. In allegato la traduzione di BIBFRAME, di Iolanda Cristalli e di Roberto Morellato, disponibile all'indirizzo: <<http://digitalia.iccu.sbn.it>>.
- Mauro Guerrini, *BIBFRAME. Un'ipotesi di ambiente bibliografico nell'era del web*, in *Il libro al centro. Percorsi fra le discipline del libro in onore di Marco Santoro*. Studi promossi da Rosa Marisa Borraccini, Alberto Petrucciani, Carmela Reale, Paola Zito, a cura di Carmela Reale. Napoli: Liguori, 2014, p. 103-115.
- Mauro Guerrini, *Classificazioni del sapere: web semantico, linked data e ontologie. Il ruolo rinnovato delle biblioteche nella trasmissione della conoscenza registrata*, in *Noetica vs informatica. Le nuove strutture della comunicazione scientifica. Atti del convegno internazionale, Roma, 19-20 novembre 2013*, a cura di Fiammetta Sabba. Firenze: Olschki, 2015, p. 145-155.
- Mauro Guerrini, *I Principi internazionali di catalogazione (ICP)*, "Biblioteche oggi", vol. 29, n. 9 (novembre 2011), p. 5-11.
- Mauro Guerrini, *Introduzione al seminario "Global interoperability and linked data in libraries"*, "Jlis.it", vol. 4, no.1 (gennaio/January 2013); special issue: *Global Interoperability and linked data in libraries*, <<http://leo.cineca.it/index.php/jlis/article/view/6307>>.
- Mauro Guerrini, *Verso nuovi principi e nuovi codici di catalogazione*, a cura di Carlo Bianchini. Milano: Sylvestre Bonnard, 2005.
- Mauro Guerrini, con Carlo Bianchini e Andrea Capaccioni, *La biblioteca spiegata agli studenti universitari*. Milano: Editrice Bibliografica, 2012.

- Mauro Guerrini, con Giuliano Genetasio, *I Principi internazionali di catalogazione (ICP). Universo bibliografico e teoria catalografica all'inizio del 21. secolo*. Milano: Editrice Bibliografica, 2012.
- Mauro Guerrini, Tiziana Possemato, *Linked data: a new alphabet for the semantic web*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cineca.it/index.php/jlis/article/view/6305>>.
- Mauro Guerrini, Tiziana Possemato, *Linked data: un nuovo alfabeto del web semantico*, "Biblioteche oggi", vol. 30, n. 3 (aprile 2012), p. 7-15, <<http://www.bibliotecheoggi.it/content/201200300701.pdf>>.
- Tom Heath, Christian Bizer, *Linked data: evolving the web into a global data space*. San Rafael, CA: Morgan & Claypool Publisher, 2011, <<http://linkeddatabook.com/editions/1.0/#coolURIs>>.
- Steffen Hennicke, Marlies Olenksy, Victor de Boer, Antoine Isaac, Jan Wielemaker, *Conversion of EAD into EDM linked data. Proceedings of the 1st International Workshop on Semantic Digital Archives (SDA 2011)*, <<http://ceur-ws.org/Vol-801/paper7.pdf>>.
- Pascal Hitzler, Krzysztof Janowicz, *Linked data, big data, and the 4th paradigm*, <<http://semantic-web-journal.net/system/files/swj488.pdf>>.
- Lynne C. Howarth, *ISBD as bibliographic content standard: interweaving threads, contemplating a future*, "Cataloging & Classification Quarterly", vol. 52, no. 8 (October 14, 2014), p. 982-999.
- Eero Hyvönen, *Publishing and using cultural heritage linked data on the semantic web*. San Rafael, CA: Morgan & Claypool Publisher, 2012.
- Antonella Iacono, *Linked data*. Roma: Associazione italiana biblioteche, 2014.
- ICCU, *Regole italiane di catalogazione. REICAT*, a cura della Commissione permanente per la revisione delle Regole italiane di catalogazione, Roma: ICCU, 2009.
- Kingsley Idehen, *Creating, deploying and exploiting linked data*, 2012, p. 35, <[http://virtuoso.openlinksw.com/presentations/Creating_Deploying_Exploiting_Linked_Data2/Creating_Deploying_Exploiting_Linked_Data2_TimBL_v3.html#\(1\)](http://virtuoso.openlinksw.com/presentations/Creating_Deploying_Exploiting_Linked_Data2/Creating_Deploying_Exploiting_Linked_Data2_TimBL_v3.html#(1))>.
- IFLA. *IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its glossary*, by Barbara Tillett & Ana Lupe Cristán (eds.). München: Saur, 2009, <<http://www.ifla.org/publications/ifla-series-on-bibliographic-control-37>>; traduzione italiana: *Dichiarazione di Principi internazionali di*

- catalogazione*. Roma: ICCU, 2009, <www.ifla.org/files/assets/cataloguing/icp/icp_2009-it.pdf>.
- IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional requirements for bibliographic records. Final report*. München: K.G. Saur, 1998, <<http://www.ifla.org/files/cataloguing/frbr/frbr.pdf>>; traduzione italiana: *Requisiti funzionali per record bibliografici*. Roma: ICCU, 2000.
- IFLA Task Force on Guidelines for OPAC Displays. *Guidelines for Online Public Access Catalogue (OPAC) displays*. The Hague, Netherlands, 2005, <<http://archive.ifla.org/VII/s13/guide/opac/guide03.pdf>>; traduzione italiana parziale pubblicata in: Andrea Marchitelli, Giovanna Frigimelica, *OPAC*, Roma: Associazione italiana biblioteche, 2012.
- IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). *Functional Requirements for Authority Data: A conceptual model. Final report, December 2008*. München: K.G. Saur, 2009.
- IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR), *Functional Requirements for Subject Authority Data (FRSAD): a conceptual model*, editors: Marcia Lei Zeng, Maja Žumer, Athena Salaba. Berlin: De Gruyter Saur, 2011, <<http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>>.
- IFLA. ISBD Review Group, and IFLA Section on Cataloguing, *ISBD, International Standard Bibliographic Description. Consolidated Edition. Recommended by the ISBD Review Group. Approved by the Standing Committee of the IFLA Cataloguing Section*. Berlin; München: De Gruyter Saur, 2011, traduzione italiana: *ISBD, International Standard Bibliographic Description, Edizione consolidata, raccomandata dall'ISBD Review Group, approvata dallo Standing Committee dell'IFLA Cataloguing Section*. Edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2012/ISBD_2012_trad_it_online.pdf>.
- Maria Grazia Iorio, Tiziana Possemato, *RDA: la catalogazione ai tempi del web*, "Biblioteche oggi", vol. 29, n. 9 (2011), p. 12-16.
- Italia. Agenzia per l'Italia digitale della Presidenza del Consiglio dei ministri, *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico (anno 2014)*, allegato alla Determinazione

- commissariale n. 95/2014 del 26-06-2014, <http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.7_finale.pdf>.
- Italia. Agenzia per l'Italia digitale della Presidenza del Consiglio dei ministri, *Linee guida per l'interoperabilità semantica attraverso i linked open data*, AGID, 2012, <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.
- Carl Jay Lagoze, *Lost identity: the assimilation of digital libraries into the web*. PhD thesis, Information Science, Cornell University, 2010, <<http://carllagoze.files.wordpress.com/2012/06/carllagoze.pdf>>.
- Joint Steering Committee for Development of RDA, *RDA: Resource Description and Access*, 2013, <<http://www.rda-jsc.org/rda.html>>.
- Joint Steering Committee for Development of RDA, *The registry! RDA vocabularies*, 2012, <<http://rdvocab.info/>>.
- Marie-Veronique Leroi, *Linked Heritage: a collaborative terminology management platform for a network of multilingual thesauri and controlled vocabularies*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/5471>>.
- Library Linked Data Incubator Group, *Library Linked Data Incubator Group final report*, 2011, <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>.
- Library of Congress. *Bibliographic Framework as a Web of Data: linked data model and supporting services*. Washington, DC: Library of Congress, 2012, <<http://www.loc.gov/bibframe/pdf/marclld-report-11-21-2012.pdf>>.
- Library of Congress. *LC RDA core elements*, 2014, <http://www.loc.gov/aba/rda/pdf/core_elements.pdf>.
- Library of Congress. Working Group on the Future of Bibliographic Control, *On the record: report of the Library of Congress Working Group on the Future of Bibliographic Control*, Washington, D.C., Library of Congress, January 9, 2008, <<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>>.
- Linked Data Community. *Linked data. Connect distributed data across the web*, 2012, <<http://linkeddata.org/>>.
- Valeria Lo Castro, *Web semantico e linked data: best practices, prospettive e criticità*, "Nuovi Annali della Scuola speciale per archivisti e bibliotecari", a. 28 (2014), p. [207]-221.
- Maurizio Lunghi, Chiara Cirinnà, Emanuele Bellini, *Trust and*

- persistence for internet resources*, “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013).
- Martin Malmsten, *Cataloguing in the open. The disintegration and distribution of the record*, “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013), p. 417-423, <<http://leo.cilea.it/index.php/jlis/article/view/5512>>.
- Andrea Marchitelli, *Il catalogo connesso*, “Biblioteche oggi”, vol. 32, n. 6 (luglio–agosto) 2014, p. 5-16.
- Giovanna Mazzola Merola, *Verso le nuove RICA*, “Bollettino AIB”, vol. 45, n. 2 (giugno 2005), p. 139-148.
- Giovanni Menduni, *Come ordinare gli open data di Firenze con l'aiuto di una biblioteca*, 2012, <<http://www.chefuturo.it/2012/07/come-abbiamo-ordinato-gli-open-data-di-firenze-con-laiuto-di-una-biblioteca/>>.
- Bertrand Meyer, *Object-Oriented software construction*. Second edition. Prentice Hall PTR. Upper Saddle River, New Jersey, 1997.
- Erik T. Mitchell, *Library linked data: research and adoption*, “Library Technology Reports”, vol. 49, no. 5 (July 2013).
- No hace falta ser matemático para ser experto en 'big data'*, <http://economia.elpais.com/economia/2015/02/12/actualidad/1423740778_503816.html>.
- Nodalities. The magazine of the semantic web*, issue 1, <https://www.zotero.org/groups/libraries_and_semantic_web/items/itemKey/BTF7TJ8K>.
- Natalya F. Noy, Deborah L. McGuinness, *Ontology development 101: a guide to creating your first ontology*, <http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html>.
- Open Knowledge Foundation, *Open data handbook*, <<http://open-datahandbook.org/it>>.
- Valdo Pasqui, *La conservazione del digitale: strategie, tecniche, standard ed esperienze per la costruzione di trusted repository di pubblicazioni periodiche*, in *Collezioni digitali di periodici in Italia ed in Europa: standard, applicazioni, valutazioni, prospettive. Atti del convegno, Arco, 15-16 novembre 2007*, a cura di Elena Corradini e Francesca Rocchetti. Trento: Giunta della Provincia autonoma di Trento, 2009, p. 41-56.
- Alberto Petrucciani, *Ragioni e principi della revisione delle RICA: per un nuovo codice italiano di catalogazione*, “Bollettino AIB”, vol. 45, n. 2 (giugno 2005), p. 149–185.

- Tiziana Possemato, *OpLiDaF: Open Linked Data Framework: una piattaforma per la creazione e la pubblicazione di linked data*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/6313>>.
- Shiyali Ramamrita Ranganathan, *Classified catalogue code, with additional rules for dictionary catalogue code*, 5th ed. Bombay: Asia Publishing House, 1964.
- Pat Riva, *FRBR Review Group initiatives and the world of linked data*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/5482>>.
- Alberto Salarelli, *Sul perché, anche nel mondo dei linked data, non possiamo rinunciare al concetto di documento*, "AIB studi", vol. 54, n. 2/3 (maggio/dicembre 2014), p. 279-293.
- Marco Santoro, con la collaborazione di Gianfranco Crupi, *Lezioni di bibliografia*. Milano: Editrice Bibliografica, 2012.
- Imma Subirats, Marcia L. Zeng, *LODE-BD Recommendations 2.0: How to select appropriate encoding strategies for producing Linked Open Data (LOD)-enabled bibliographic data*. Rome: Food and Agriculture Organization of United Nations, updated 2013-07-28, <<http://aims.fao.org/lode/bd>>.
- Elaine Svenonius, *The Intellectual Foundation of Information Organization*, Cambridge, Mass.: MIT Press, 2000; traduzione italiana: Elaine Svenonius, *Il fondamento intellettuale dell'organizzazione dell'informazione*, a cura di Maria Letizia Fabbrini, Firenze: Le Lettere, 2008.
- Teaching and learning, connected*, <<http://www.talis.com/>>.
- Roy Tennant, *MARC must die*, "Library Journal", October 15, 2002, <<http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>>.
- Henry S. Thompson, David Orchard, *Urns, namespaces, and registries*, <<http://www.w3.org/2001/tag/doc/URNsAndRegistries-50-2006-08-17>>.
- Barbara B. Tillett, *The Bibliographic universe and the new IFLA Cataloging Principles : lectio magistralis in library science: Florence, Italy, Florence University, March 14, 2008 = L'universo bibliografico e i nuovi Principi di Catalogazione dell'IFLA : Lectio magistralis di biblioteconomia: Firenze, Università degli studi di Firenze, 14 marzo 2008 / di Barbara Tillett; traduzione di Carlo Bianchini*. Fiesole (Firenze): Casalini Libri, 2008, <<http://digital.casalini.it/9788885297814>>.

- Barbara B. Tillett, *Cataloging principles and RDA: Resource Description and Access* (webcast). Washington, D.C., Library of Congress, June 10, 2008, <http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=4327>.
- Barbara B. Tillett, *Mantenere il ruolo delle biblioteche nel web semantico tramite RDA: Resource Description and Access*, "Jlis.it", vol. 2, no. 2 (luglio/July 2011), p. 1-19, <<http://leo.cineca.it/index.php/jlis/article/view/4760>>.
- Barbara B. Tillett, *Prefazione*, in: Carlo Bianchini, Mauro Guerrini, *RDA: Resource Description and Access. Linee guida per identificare e collegare entità nel web semantico*. Milano: Editrice Bibliografica, 2014.
- Barbara B. Tillett, *RDA (Resource Description and Access): lo sviluppo di un nuovo codice di catalogazione internazionale*, "Bibliotime. Rivista elettronica per le biblioteche", vol. 11, n. 1 (2007), <<http://www2.spbo.unibo.it/bibliotime/num-xi-1/tillett.htm>>.
- Barbara B. Tillett, *RDA and the semantic web, linked data environment*, "Jlis.it", vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/6303>>.
- Barbara B. Tillett, *Resource Description and Access: background/overview* (webcast). Washington, D.C., Library of Congress, May 14, 2008, <http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=4320>.
- Seth Van Hooland, Ruben Verborgh, *Linked data for libraries, archives and museums. How to clean, link and publish your metadata*. London: Facet Publishing, 2014.
- Daniel Van Spanje, *Rise, fall of the cataloguer's empire: a changing landscape*, relazione presentata al convegno *Faster, Smarter, Richer. Reshaping the library catalogue*, Roma (Italy), 27-28 February 2014, <<http://eprints.rclis.org/22835/>>.
- Daniel Vila Suero, Elena Escolano Rodríguez, *Linked data at the Spanish National Library and the application of IFLA RDFS models*. "SCATNews", no. 35 (June 2011), <<http://www.ifla.org/files/assets/cataloguing/scatn/scat-news-35.pdf>>.
- Boris Villazón-Terrazas, Luis. M. Vilches-Blázquez, Oscar Corcho e, Asunción Gómez-Pérezin, *Methodological guidelines for publishing linked data* del 2011, <https://www.lri.fr/~hamdi/datalift/tuto_inspire_2012/Suggestedreadings/egovld.pdf>.
- Julius Volz (e altri), *Discovering and maintaining links on the web of data*. Berlin; Heidelberg: Springer 2009.

- Norman Walsh, *Names and addresses*, volume 9, issue 73 (25 Jul 2006) (modified 08 Oct 2010), <<http://norman.walsh.name/2006/07/25/namesAndAddresses>>.
- Romain Wenz, *Linked open data for new library services: the example of data.bnf.fr*, “Jlis.it”, vol. 4, no. 1 (gennaio/January 2013), <<http://leo.cilea.it/index.php/jlis/article/view/5509/>>.
- Paul Gabriele Weston, *Catalogazione bibliografica. Dal formato MARC a FRBR*, “Bollettino AIB”, vol. 41, n. 3 (settembre 2001), p. 267-283. <<http://bollettino.aib.it/article/view/8536/7616>>.
- Mirna Willer, Gordon Dunsire, *Bibliographic information organization in the semantic web*. Oxford, Cambridge; New Delhi: Chandos Publishing, 2013.
- Mirna Willer, Gordon Dunsire, *ISBD, the UNIMARC Bibliographic Format, and RDA: interoperability issues in namespaces and the linked data environment*, “Cataloging & Classification Quarterly”, vol. 52, no. 8 (2014), p. 888-913.
- Mirna Willer, Gordon Dunsire, Boris Bosančić, *ISBD and the semantic web*, “Jlis.it”, vol. 1, no. 2 (dicembre/December 2010), p. 213-236, <<http://leo.cineca.it/index.php/jlis/article/view/4536/4408>>.
- Paul Walk’s weblog, *Linked, open, semantic?*, <<http://blog.paulwalk.net/2009/11/11/linked-open-semantic/>>.
- World Wide Web Consortium. Incubator Group, *Library Linked Data Incubator Group final report*, October 25, 2011, <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025>>.
- World Wide Web Consortium. *W3C Semantic web activity*, <www.w3.org/2001/sw>.

INDICE DEI NOMI

Sono esclusi i nomi citati nella
bibliografia finale*

@Cult, 15, 18, 145, 148, 158
 Agenda digitale europea, 20, 123
 Agenzia per l'Italia digitale, 103,
 109, 121, 122
 AIB, 15, 75, 93, 133, 181, 183,
 184, 185, 191
 AIB Lazio, 16
 AIB Toscana, 133
 AIB Trentino-Alto Adige, 16
 American Library Association
 (ALA), 31, 32
 Anibaldi Stefano, 214, 215
 Artium- Centro museo Vasco de
 Arte Contemporaneo, 148, 158
 Associazione italiana biblioteche
 (AIB), 15, 75, 93, 133, 181,
 183, 184, 185, 191
 Baker Thomas, 16
 Bellini Emanuele 17, 68
 Belpassi Eleonora, 17
 Bergamin Giovanni, 66
 Berners Lee Tim, 23, 24, 28, 30,
 31, 33, 34, 38, 55, 56, 59, 71,
 82, 93, 94, 100, 101, 112, 147,
 185, 197
 Bianchini Carlo, 19, 20, 65, 169,
 171, 180, 181, 182, 184, 185,
 186, 188, 191, 192, 194, 205
 Biblioteca Angelica di Roma, 174

Biblioteca Nazionale Centrale di
 Firenze (BNCF), 15, 70, 179,
 209, 210
 Bibliotheque nationale de France
 (BnF), 64, 70, 85, 149, 150,
 151, 152, 154, 157, 203, 236
 Bizer Christian, 28, 31, 38, 41,
 56, 241, 246
 Borraccini Rosa Marisa, 134
 British Library, 86, 152
 Bosančić Boris, 225
 Bucci Helena, 13
 Capaccioni Andrea, 171
 Caproni Attilio Mauro, 169
 Casalini libri, 15, 133, 184, 187
 Celli Fabrizio, 214, 215
 CERN (European Organization
 for Nuclear Research,
 organizzazione europea per la
 ricerca nucleare), 23
 Cirinnà Chiara, 68
 Comune di Firenze, 15, 200
 Conferenza dei rettori delle
 università italiane (CRUI), 15
 Corcho Oscar, 106
 Corradini Elena, 68
 Coyle Karen, 13, 32, 78, 84, 172,
 174, 195, 204
 Cristalli Iolanda, 134
 Cristàn Ana lupe, 132, 183
 Crupi Gianfranco, 17, 58, 173,
 200
 Cyganiak Richard, 55, 112, 113

* L'indice è redatto da Daniela Giglio

Davies Tim, 96
 De Pinedo Isa, 184
 De Robbio Antonella, 75, 113, 197, 198
 Decker Stefan, 199
 Delle Donne Roberto, 19, 168
 Deutsches Nationalbibliothek (DNB), 155
 Digital Preservation Department degli UK National Archives, 90
 Doerr Martin, 213
 Dublin Core Metadata Initiative (DCMI), 84, 86, 152
 Dunsire Gordon, 15, 24, 86, 187, 194, 198, 201, 224, 225, 249
 Escolano Rodríguez Elena, 187, 249
 Europeana, 101, 116, 213
 Fabbrini Maria Letizia, 171
 Fischetti Mark, 23
 Folch Maria, 214
 Fondazione Rinascimento Digitale, 15, 68
 Frie Universität zu Berlin, 111, 224
 Frigimelica Giovanna, 183
 Genetasio Giuliano, 132, 144, 186
 Giglio Daniela, 17
 Giordani Anna, 42
 Gómez-Pérez Asunción, 106, 249
 Gradmann Stefan, 213
 Guerrini Mauro, 15, 16, 17, 19, 20, 21, 58, 60, 65, 132, 134, 144, 171, 180, 186, 191, 192, 196, 200, 205
 Heath Tom, 28, 31, 38, 41, 56
 Hendler James, 82
 Hennicke Steffen, 213
 Hillmann Diane, 201
 HM Treasury, 20
 Howarth Lynne C., 194
 Hungarian National Library, 116
 Hyland Bernadette, 107
 Idehe Kingsley, 202
 IFLA, 65, 85, 86, 87, 132, 137, 144, 155, 181, 182, 183, 184, 186, 188, 198, 226
 IFLA Section on Cataloguing, 186
 IFLA Study Group on the Functional Requirements for Bibliographic Record, 137, 181
 IFLA Working Group on Functional Requirements and Numbering of Authority records, 82, 182
 IFLA. ISBD Review Group, 186
 International Federation of Library Association and Institutions (IFLA), 85, 187
 Isaac Antoine, 213
 Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), 15, 37, 38, 86, 132, 133, 134, 149, 181, 184, 186, 200
 Istituto di teoria e tecniche dell'informazione giuridica del Consiglio nazionale delle ricerche (ITTIG.CNR), 15
 Jaques Yves, 214, 215
 Jarrar Mustafa, 46
 Jentzsch Anja, 55, 113
 JLIS.it, 25, 58, 66, 149, 182, 186, 200, 204, 225
 Joint Steering Committee for

Development of RDA, 65, 86,
 185, 186, 188
 Keizer Joannes, 214, 215
 Lagoze Carl Jay, 213
 Lassilla Ora, 82
 Libera Università di Bolzano, 16
 Library Linked Data Incubator
 Group, 14, 64, 79, 86, 134,
 166, 179, 185, 198, 199, 215,
 237
 Library of Congress, 36, 37, 70,
 87, 126, 133, 134, 142, 151,
 154, 179, 194, 195, 202, 209,
 210, 212
 Library of Congress Working
 Group on the Future of
 Bibliographic Control, 126
 Linked Data Community, 29
 Linking Open Data Community
 (W3C SWEO), 111
 Lucarelli Anna, 66
 Lunghi Maurizio, 68
 Maiello Rosa, 16
 Malmsten Martin, 204
 Mancini Mario, 17
 MARC of Quality (TMQ), 140
 Marchitelli Andrea, 17, 183
 Mazzola Merola Giovanna, 184
 Max-Planck-Gesellschaft, 20
 Meghini Carlo, 213
 Menduni Giovanni, 200
 Montiel-Ponsoda Elena, 249
 Morellato Roberto, 134
 Nowack Benjamin, 97
 OCLC (Online Computer
 Library Center), 15, 68, 90,
 194, 208, 209, 212
 Open Link Software (UK), 111
 Orchard David, 31
 Pasqui Valdo, 16, 68
 Peruginelli Ginevra, 17, 58, 200
 Petrucciani Alberto, 184
 Phipps Jon, 201
 Possemato Tiziana, 16, 19, 20,
 21, 58, 60, 186, 196
 Presidenza del Consiglio dei
 Ministri, 103, 109, 121
 Provincia Autonoma di Trento,
 20, 68, 91
 Ranganathan Shiyali Ramamrita,
 182, 193
 Reale Carmela, 134
 Research Libraries Group
 (RLG), 90
 Rocchetti Francesca, 68
 Rossi Gianni, 41, 42, 59, 63
 Salaba Athena, 182
 Santoro Marco, 173
 Sapienza Università di Roma, 16
 Saur K.G., 85, 86, 132, 137, 181,
 182, 183, 186
 Sabba Fiammetta, 34
 SBN (servizio bibliotecario
 nazionale) 37, 70, 148, 149,
 174, 180, 209
 ScanBit, 148, 158
 Schreiber Guus, 213
 Serrai Alfredo, 169
 Solodovnik Iryna, 16, 73, 207
 Sompel Herbert van de, 213
 Stellato Armando, 214
 Subirats Imma, 73
 Svenonius Elaine, 171, 180, 182,
 191
 Szépművészeti Múzeum, 148,
 158
 Tecnalìa, 148, 158
 Tennant Roy, 132, 181

Thompson Henry S., 31	Vila Suero Daniel, 249
Tillet Barbara B., 15, 132, 183, 184, 186, 192	Vilches-Blázquez Luis. M., 106
Università di Firenze, 13, 15, 16, 63, 187, 200	Villazón-Terrazas Boris, 106, 249
Università di Napoli Federico II, 16, 20	Viti Elisabetta, 17, 210
Universität Leipzig, 111	Walk Paul, 102
University of Oregon Libraries, 165	Walsh Norman, 31
Van Hooland Seth, 14, 204	Wenz Romain, 149
Van Spanje Daniel, 192	Weston Paul Gabriele, 181
VAST LAB – PIN, 149	Willer Mirna 24, 186, 187, 194, 198, 224, 225
Verborgh Ruben, 14, 204	Zeng Marcia Lei, 73, 182
	Zito Paola, 134
	Žumer Maja, 182

